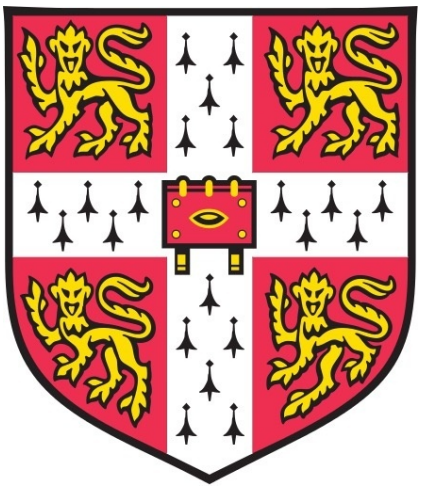
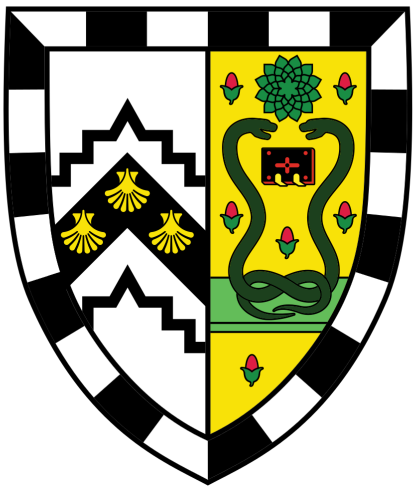
Characterization of ToxIN$_{Pa}$ sensitive bacteriophages of *Serratia* sp. ATCC 39006

Bihe Chen





Department of Biochemistry
Gonville & Caius College, University of Cambridge

September 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Characterization of ToxIN$_{Pa}$-sensitive bacteriophages of *Serratia* sp. ATCC 39006

Bihe Chen

Summary

Abortive infection (Abi) is an anti-phage mechanism in which a bacterium initiates its own death upon phage infection. This prevents or decreases the production of phage progeny and protects clonal siblings in the bacterial population by an act akin to an "altruistic suicide." Abortive infection can be mediated by a Type III toxin-antitoxin system called ToxIN$_{Pa}$, first found in *Pectobacterium atrosepticum*. ToxIN$_{Pa}$ consists of an endoribonuclease toxin and RNA antitoxin and is inactive as a heterohexameric complex until infection by certain phages causes destabilization of ToxIN$_{Pa}$, leading to bacteriostasis and, eventually, lethality. However, it is still unknown why only certain phages are able to activate ToxIN$_{Pa}$-mediated Abi and what mechanisms are involved.

This study aimed to address this issue by first introducing ToxIN$_{Pa}$ into the Gram-negative enterobacterium, *Serratia* sp. ATCC 39006 (*S*39006). A novel environmental *S*39006 phage, ΦCBH8, that is sensitive to ToxIN$_{Pa}$-mediated Abi was isolated, characterized, and compared to its spontaneous "escape" mutants that were insensitive to ToxIN$_{Pa}$. Characterization of ΦCBH8 led to the discovery of a new genus of T4-family phages that are all ToxIN$_{Pa}$-sensitive. Genomic comparison of ΦCBH8 with its spontaneous ToxIN$_{Pa}$-escape mutants led to the discovery of two distinct genetic loci that are candidates for activation of ToxIN$_{Pa}$-mediated Abi.

One such locus was a multi-gene region of the phage genome, the deletion of which enabled some ΦCBH8 mutants to escape ToxIN$_{Pa}$. Different ToxIN$_{Pa}$-escape mutants had variable deletions in their genomes, but this multi-gene locus was deleted in all of the deletion mutants. Experiments showed that at least two genes in this locus must be simultaneously involved in activating ToxIN$_{Pa}$-mediated Abi and one phage gene product within the locus (ORF18) was extremely toxic to the bacterial host.

Another locus was the *asiA* gene. This study showed that ΦCBH8 mutants with mutated versions of *asiA* became insensitive to ToxIN$_{Pa}$ but genetic complementation restored their sensitivity. AsiA is involved in σ-appropriation but the mutated AsiA in ΦCBH8 mutants was no longer functional. Further experiments showed that AsiA from wild type ΦCBH8 reduced ToxIN$_{Pa}$ promoter activity while mutant AsiA did not, suggesting that AsiA could be involved in activating ToxIN$_{Pa}$-mediated Abi by perturbing *toxIN$_{Pa}$* transcription.

Finally, an outer membrane protein (OmpW) was defined as the ΦCBH8 receptor in *S*39006. Transfer of the *S*39006 *ompW* gene to other bacterial species rendered some of them sensitive to the phage thereby enabling expansion of ΦCBH8 host range. Results showed that ΦCBH8 was insensitive to ToxIN$_{Pa}$ in *E.coli,* suggesting that activation of ToxIN$_{Pa}$-mediated Abi is host-dependent. The function of OmpW was explored to try to link bacterial host physiology with phage susceptibility

In conclusion, this work characterized novel phage genes involved in activating ToxIN$_{Pa}$-mediated Abi. This provided evidence that viral gene lethality and σ-appropriation might be related to the mechanism(s) for phage activation of ToxIN$_{Pa}$-mediated Abi.

Declaration

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

The length of this thesis does not exceed the limit set by the Biology degree committee.

This research project was performed in the laboratory of Professor George Salmond, Department of Biochemistry at the University of Cambridge, except for specific elements of the work that were performed elsewhere as described herein, during the period September 2014 to September 2018.

Bihe Chen
September 2018

Acknowledgements

I am grateful to Professor George Salmond for the opportunity to carry out this interesting project in the Biochemistry Department where many discoveries that changed the course of human history were born. I thank George for his continued support and guidance throughout my Ph.D. project. His many sparks of genius often led to another new idea and one wishing that life could be infinite to forever be doing Ph.Ds. My special thanks also go to Dr. Martin Welch, whose humor and candidness is often thought-provoking.

I am also thankful to the Cambridge Trust for funding my course and making it possible to study in Cambridge. Without the generous financial aid from the prestigious scholarship, I would be shivering in The Windy City now and never would have known the "Philosophy" aspect of the Ph.D. that is deeply rooted in 800 years of history. I am also thankful to the Cambridge Philosophical Society and Caius College for supporting my trips abroad where I have learnt that science extends beyond borders.

Of course, the Ph.D. life will not be complete without all the lovely faces in the Salmond Lab. My gratitude goes to those who have dedicated so much time to making the Lab a safe and effective place to conduct research - Ms. Alison Rawlinson, Ms. Diana Breitmaier, Dr. Rita Monson and Dr. Jessica Bergman. I am also lucky to have had the support of other colleagues in the lab, especially those who started in the same year as me, while we treaded in uncharted waters. The Salmond Lab is also a place where great friendships blossom, and I owe it to Svenja, Nathalie, Tracy and Dr. Lee for all the joy. There are also people outside of the lab without whom I would not be who I am today. Thank you to all my friends in college and in the CGS, and those who never failed to accompany me to Starbucks. Above all else, my deepest thanks and love to Maria, Yini and Paula. The three of you have taught me the meaning of kindness, love and faith.

Finally, and most importantly, no words could describe my gratitude to my parents. Mom and Dad, you have been with me from the beginning to the end, through thick and thin, and never stopped believing in me. Thank you so much and I love you both.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abi	abortive infection
Ap	ampicillin
BHL	N-butanoyl-L-homoserine lactone
BREX	bacteriophage exclusion
Cas	CRISPR-associated genes
cfu	colony forming unit
Cm	chloramphenicol
CRISPR	clustered regularly interspaced short palindromic repeats
CTD	C-terminal domain
DAPA	2,6-Diaminopimelic acid
dH_2O	deionized H_2O
dNTP	deoxynucleotide
dsDNA	double stranded DNA
dsRNA	double stranded RNA
EDTA	ethylenediaminetetraacetic acid
EOP	efficiency of plating
EtOH	ethanol
Glu	D-Glucose
ICTV	International Committee on Taxonomy of Viruses
Kb	kilobase
Kn	kanamycin
L-ara	L-arabinose
LB(A)	Luria broth (agar)
LPS	lipopolysaccharides
MCP	major capsid protein
MES	2-(N-morpholino) ethanesulfonic acid
MIC	minimum inhibitory concentration
MOI	multiplicity of infection
MUG	4-Methylumbelliferyl beta-D-galactoside
NTD	N-terminal domain

OD_{600}	optical density measured at a wavelength of 600 nm
OMP	outer membrane protein
ORF	open reading frame
PBS	phosphate-buffered saline
PCR	Polymerase chain reaction
pfu	plaque forming unit
PLG	phase-lock gel
ppGpp	tetra- and penta- guanosine phosphates
PTA	phosphotungstic acid
R-M	restriction-modification
RNAP	RNA polymerase
RLU	relative light units
rpm	revolutions per minute
*S*39006	*Serrtia* sp. ATCC 39006
SaPIs	staphylococcal pathogenicity islands
SDS	sodium dodecyl sulfate
Sie	superinfection exclusion
ssDNA	single stranded DNA
TA	toxin-antitoxin
TAE	Tris-acetate-EDTA
Tc	tetracycline
TEM	transmission electron microscopy
TSA	tryptone swarm agar

List of Publications

This work has so far contributed to two articles in peer reviewed journals, both of which can be found in the Appendix IV of this thesis:

1. Chen B, Akusobi C, Fang X, Salmond G. Environmental T4-Family Bacteriophages Evolve to Escape Abortive Infection via Multiple Routes in a Bacterial Host Employing "Altruistic Suicide" through Type III Toxin-Antitoxin Systems. Frontiers in Microbiology. 2017;8.

2. Goeders N, Chai R, Chen B, Day A, Salmond G. Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems. Toxins. 2016;8(10):282.

Chapter 1- Introduction

1. Bacteriophages

1.1 The discovery of bacteriophages and their abundance

Bacteriophages (phages) are viruses that infect bacteria. As obligate parasites of bacteria, phages are believed to have been first discovered by the English bacteriologist Frederick Twort in 1915 when he observed the "glassy transformation" of micrococci during cultivation (1). However, at that time Twort could not confirm the nature of the glassy transparent material he observed, which we now know to be zones of dead bacteria formed after phage propagation. Shortly after, in 1917, the French-Canadian biologist Felix d'Herelle observed clear spots speckling the cultures of coccobacillus grown on agar and believed that they were caused by obligate parasites of bacteria (2). Felix d'Herelle's arguably independent discovery finally led to the acknowledgement of the existence of phages and marked the official beginning of a century's research on phages. In fact, it was d'Herelle who first used the term "bacteriophage".

Today, phages are considered to be the most abundant biological entity in the world, with an estimated number of 10^{31} viral particles in the biosphere (3). Perhaps what is most impressive is the diversity of biology they represent. Phages have been found in every environment where bacteria live. At the same time, different phages are distinct from each other both genetically and morphologically with genomes varying from 3.4 kb to almost 500 kb and morphologies ranging from tailed phages to their non-tailed counterparts (4). However, our knowledge of phages is still very limited. In 2008 there were only 500 phage genomes deposited in GenBank (3). That number has grown to more than 5,000 today but they still only represent a tiny fraction of the entire phage population, most of which are not yet isolated or sequenced. Even for the deposited phage genomes in GenBank, little is known about the functions of many phage genes. Indeed, phages contain the largest reservoir of unexplored genetic information on earth.

1.2 Phage classification and structure

1.2.1 Conventional classification of phages based on morphology

To better characterize phages and address their diversity, classification is necessary. However, no universal phage classification method exists, partly because, unlike their bacterial hosts, phages

lack universal genes that can be used to identify phylogenetic relationships (5). Therefore, the official phage taxonomy published by the International Committee on Taxonomy of Viruses (ICTV) lags far behind the discovery of novel phage types (6). So far, the most commonly used phage classification method is based on David Bradley's proposal in 1967 that focuses on the morphological characteristics of phages (7). This divides phages into "tailed phages" (*Caudovirales*) and "non-tailed phages" that includes "cubic phages", "filamentous phages" and "pleomorphic phages" (Figure 1.1).

Tailed phages are the most abundant type. More than 96% of phages are tailed phages and all of them have a double stranded DNA (dsDNA) genome contained within their icosahedral or elongated heads. They have no envelop and adsorb to their bacterial host directly from the outside (6). Tailed phages can be further divided into 3 families based on their tail properties: phages in the *Myoviridae* family have long, contractile tails; phages in the *Siphoviridae* family have long, non-contractile tails and represent the largest family of tailed phages; phages in the *Podoviridae* family have short, non-contractile tails (6, 8). The *Myoviridae* family includes some of the most well-studied phages such as T4, the structure of which will be elaborated in Section 1.2.2.

Non-tailed phages are less common but have also been isolated against a wide variety of bacterial hosts. "Cubic phages" include families such as *Microviridae* (ssDNA), *Corticoviridae* (dsDNA), *Tectiviridae* (dsDNA), *Leviviridae* (ssRNA)… *etc*. Some "cubic phages" possess a membrane envelop that protects their capsid but all of them lack a permanent phage tail. "Filamentous phages" also include several families, the most well-known of which is the *Inoviridae* (ssDNA) composed of long and rod-shaped phages. Some members of the *Inoviridae* can integrate into the genomes of bacterial hosts and contributes to the production of major virulence factors in *Vibrio cholerae*. "Pleomorphic phages" include seven different families all with dsDNA genomes but are less commonly found than other phage families. "Pleomorphic phages" include a wide variety of irregular morphologies and some families don't even possess capids (6) (Figure 1.1).

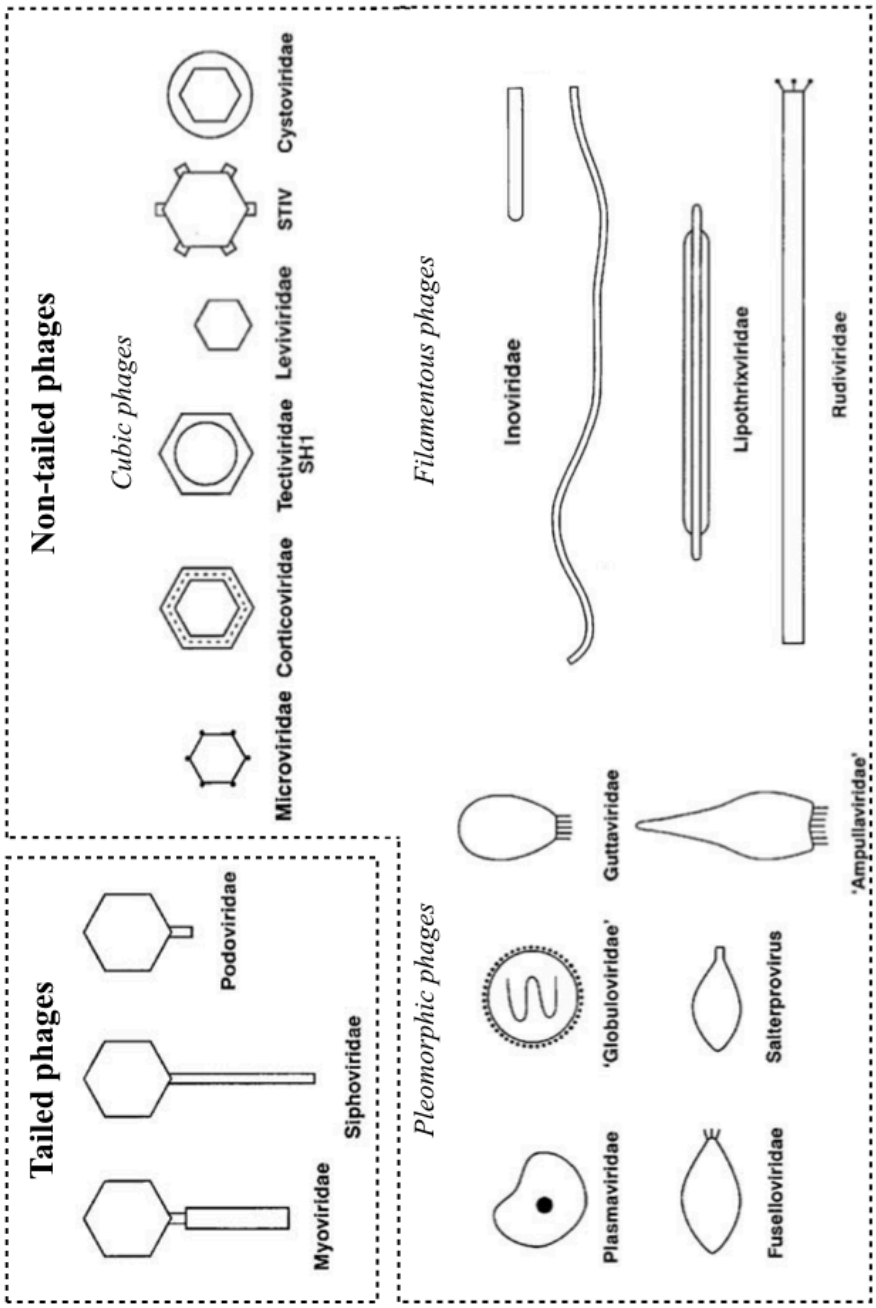


Figure 1.1: Classification of phages based on morphology. This figure is reproduced from the book chapter published by Ackermann and colleagues (9). Schematic illustrations of phage morphology are shown with the names of phage families. Phages are broadly categorized into "tailed" and "non-tailed", with "tailed" phages representing the most abundant and well-studied category that includes phages with long, contractile tails (*Myoviridae*), phages with long, noncontractile tail (*Siphoviridae*) and phages with short, noncontractile tail (*Podoviridae*) (9).

1.2.2 Structure of phage T4: the most well-studied member of the Myoviridae

Even though phages possess a wide range of morphological structures, past research has been focused only on a few model phages, most of them belonging to the *Caudovirales* order, such as phage T4 (*Myoviridae*), phage lambda (*Siphoviridae*) and phage T7 (*Podoviridae*). Among them, *Escherichia coli* (*E.coli*) phage T4 is the most well-studied and is a good representative of the structural complexity of *Myoviridae*.

The structure of phage T4 has been elucidated using techniques such as complementation assays, cross-linking analyses, x-ray crystallography, TEM and cryoEM (10). The main components of phage T4 structure are head, tail and long tail fibers, all of which are assembled individually before coming together to form a mature virion (Figure 1.2) (10). The prolate head of phage T4 is 1150 Å-long, 850 Å-wide with hemi-icosahedral ends and made up of 3000 polypeptide chains of at least 12 kinds of proteins (11). The phage head contains its 172-kb dsDNA genome and is connected to the phage tail. The T4 tail, 1000 Å in length and 210 Å in width, consists of an inner cylinder called the tail tube and an outer cylinder called the tail sheath that is wrapped tightly around the tail tube. While one end of the tail is connected to the phage head, the other end is associated with a hexagonal structure called the baseplate, upon which six long tail fibers are attached. Each long tail fiber of phage T4 consists of a rigid proximal part hinged with a distal part. The long tail fibers act as antennas that sense the environment and attach to receptors on bacterial surfaces. This is followed by unfolding of short tail fibers and irreversible attachment of the star-shaped baseplate to the bacteria host (10, 11). This initiates the infection process during which the tail sheath contracts and genomic DNA from the phage head will pass through the tail tube and into the host cell. Details of the infection process can be found in Section 1.3.

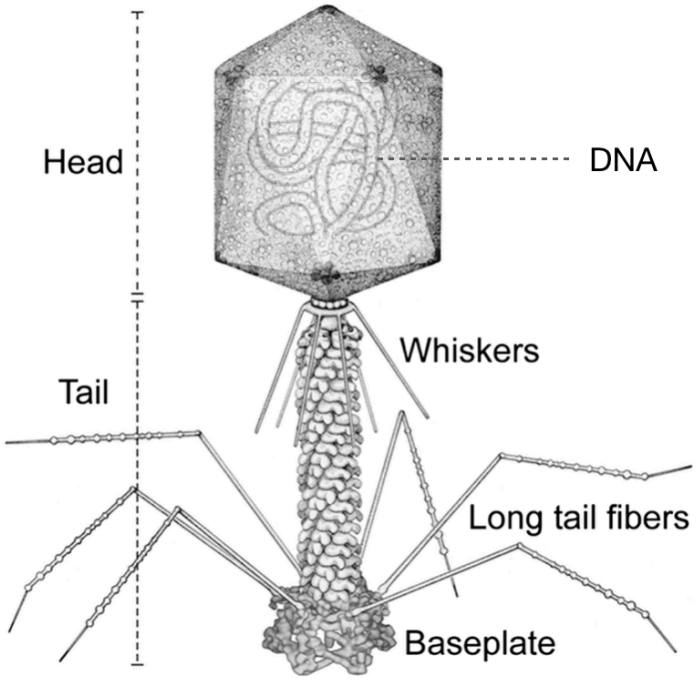


Figure 1.2: Schematic representation of phage T4 structure. The main components of phage T4 structure - head, tail, long tail fibers and baseplate are indicated in the diagram. The phage genomic DNA is shown as contained within the head capsid and enters the host through the tail tube. This figure is reproduced from the review by Yap and colleagues (10).

1.2.3 New developments in phage classification

Whilst conventional phage classification during the 1970s-1990s focused on biological attributes of phages, in particular morphological structure, recent assignment of phage taxonomy has seen new trends both in the availability of phage sequence information and parameters used to classify phages using this information.

The first new trend of phage classification is the increasing reliance on phage genomic sequence data, instead of simple phage morphology. In recent years the ICTV has attempted to use predicted proteome similarity to define genera and subfamily, as well as EMBOSS Stretcher to calculate similarities between phage genomes, yet these methods suffered from limitations. Recent phage classification practice has attempted to use alternative methods such as BLASTn to perform identification of related phages. BLASTn has been found to be superior to previous programs such as EMBOSS Stretcher and some of the newest phylogenetic analysis tools such as CLANS and mVISTA are based on BLASTn. As a result of the new genome-based phage taxonomy principle and the great increase in the number of sequenced phage genomes, today the order *Caudovirales* includes six subfamilies, 80 genera, and 441 species, compared to only 3 families, 18 genera and 36 species in 2008 (12).

The second trend of phage classification is the acceptance of metagenomics data to be used for phage taxonomy. Historically, the classification of new viruses by the ICTV required substantial biological information such as host range, replication cycle, virion structure… *etc*. However, with the availability of high-throughput sequencing and metagenomics data, phage genomes have been discovered that are not directly correlated with any biological agent. After several workshops to discuss the validity of such metagenomics data, the ICTV has decided to accept, with caution, that the detection of phage genome sequence in a sample is sufficient to infer the existence of a corresponding phage (5). Using metagenomics data, virologists have uncovered an abundance of phages in unexpected sources. For example, a dsDNA phage named crAssphage with a genome size of 97 kb was discovered from metagenomics of waste water and is six times more abundant in public metagenomics data than all other known phages combined (13).

1.3 Phage life cycle

The process through which phage infection occurs leading to phage reproduction is called the phage life cycle. Every second there is an estimated 10^{25} phage infections occurring in the biosphere (14) and the length of a phage life cycle can vary greatly. The most well-defined phage life cycles are illustrated in Figure 1.3 and can be categorized into the lytic life cycle and the lysogenic life cycle. Phages adopting either life cycle share common initial steps in the infection process such as adsorption and DNA injection but adopt different post-injection survival and replication strategies. In recent years, this bifurcation of phage lifestyle is challenged by the discovery of "pseudolysogeny" – an alternative phage life cycle where phage development inside the host cell is stalled without either genome replication and host cell lysis (as in the lytic life cycle) or genome replication and stable maintenance synchronized with the host cell cycle (as in the lysogenic life cycle). Pseudolysogeny is usually caused by stringent growth conditions for the host, such as starvation, and terminated by the availability of nutrients, after which the phage goes into either the lytic or lysogenic life cycle (15). However, the following introduction of phage life cycle is still focused on the classical lytic/lysogenic categorization, especially the lytic life cycle, because of limited understanding of pseudolysogeny and because most phages mentioned in this study adopt the lytic life cycle.

1.3.1 Adsorption and DNA injection

The first step of a phage life cycle is adsorption - the recognition and attachment to the phage receptor on the bacterial surface. Adsorption is crucial, not only because it is the first point of contact between phage and host, but also because the specificity of receptor recognition is a major contributor to determining phage host range. Taking phage T4 as an example, adsorption consists of three main steps: initial contact, reversible binding and irreversible attachment (16). Initial contact is made by random collision of the phage and host, after which the phage binds reversibly to the host surface via its long tail fibers to further its search for specific receptors. After a receptor is found, the baseplate is anchored to the receptor with the help of short tail fibers and irreversible binding is achieved (11). This allows for subsequent injection of phage DNA.

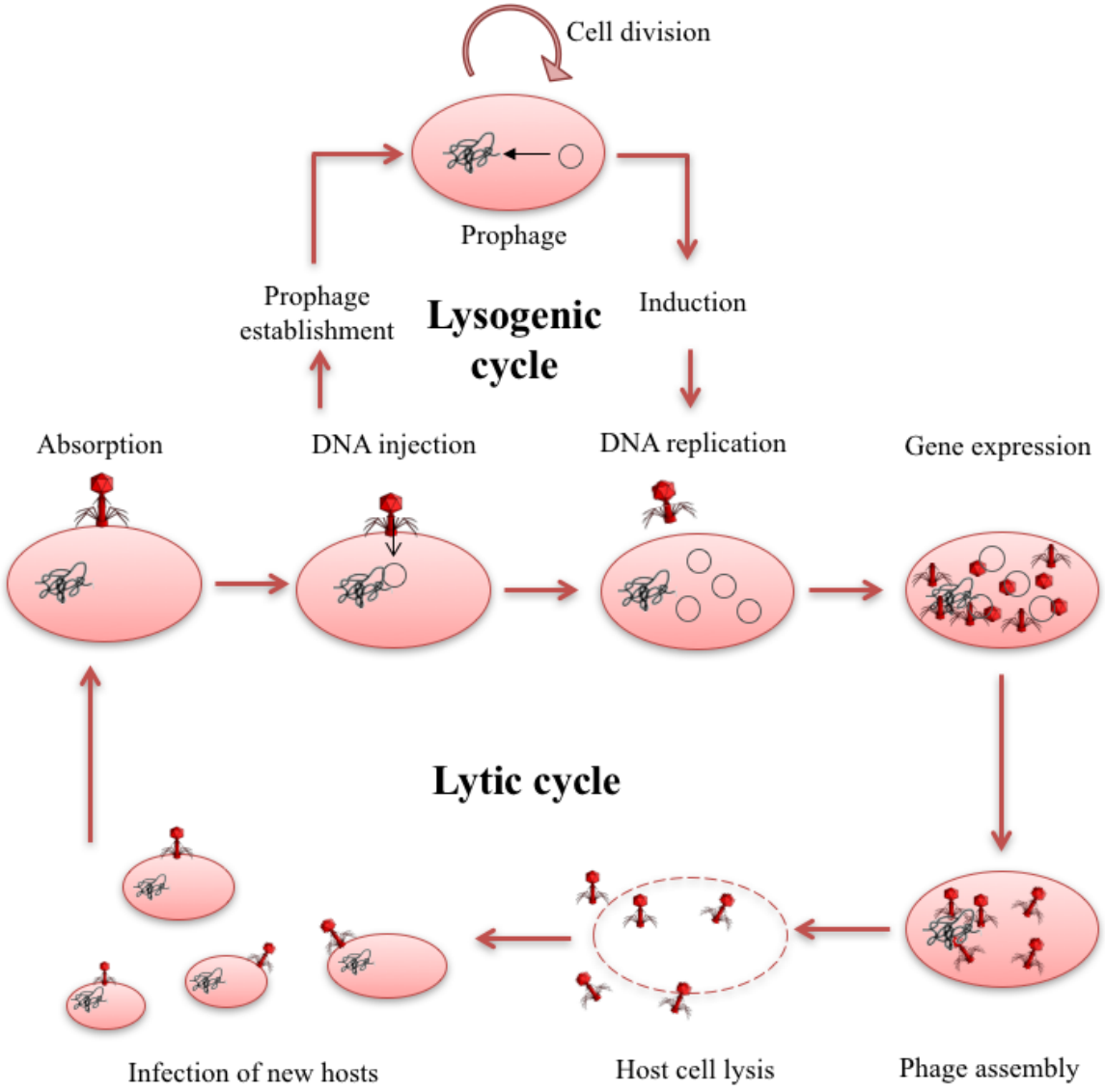


Figure 1.3: Schematic representation of phage life cycle. Phage infection is initiated with phage adsorption to the host surface, followed by injection of phage DNA into the host cytoplasm. After DNA injection, the virulent phages enter the lytic cycle while temperate phages could either enter the lysogenic cycle or the lytic cycle. In the lytic cycle, phage DNA is replicated, phage genes are expressed, and phage components are assembled into mature phage particles. This eventually leads to the lysis of the host cell and release of varying numbers of phage progenies. The progenies could restart the life cycle by adsorbing to a new host. In contrast, temperate phage entering the lysogenic cycle don't immediate begin DNA replication and gene expression. Instead, genomic DNA is circularized and integrated into specific sites in the host chromosome or maintained in the host cytoplasm as a circular replicon until exposure to certain signals (such as radiation or mutagenic

chemicals), after which the prophage genome can be excised from the host chromosome and the phage would enter the lytic cycle (27, 28).

The identity of bacteria cell surface components that phages recognizes also varies greatly, such as wall components (Gram-positive and Gram-negative bacteria), components of bacterial capsules, membrane proteins and appendages such as pili and flagella. A summary of different cell surface components that have been found to act as receptors and their specific examples are shown in Table 1.1. This diversity in receptor identity is reflected by the diversity in phage adsorption mechanisms, as reviewed in depth by Silva and colleagues (16). While most phages recognize a single receptor, in some cases phages have dual receptors or recognize different cell surface components for reversible and irreversible binding. For example, *E.coli* phage T5 utilizes the O-antigen of host lipopolysaccharide (LPS) for reversible binding and outer membrane protein FhuA for irreversible binding. In such cases, the utilization of more-exposed cell wall components for reversible binding creates higher chances of interaction and more stable interaction, which allows easier access to receptors for irreversible binding.

For tailed phages, after irreversible binding of the phage tail with host receptor, the tail tube penetrates through the bacterial surface. In this process, enzymes associated with the phage tail help penetration by digesting the peptidoglycan layer. After the phage tail has penetrated through the inner membrane of bacteria, genomic DNA is released from the phage head and injected into the host cytoplasm through the tail tube.

1.3.2 DNA replication and gene expression

Based on post-DNA injection survival strategies, phages can be divided into two categories: virulent phages that enter the lytic life cycle and temperate phages that may go through a lytic life cycle or can enter the lysogenic life cycle.

In the lytic life cycle, after DNA injection phages immediately embark on the process of making large amounts of viral DNA/RNA and proteins that eventually package into mature phage particles. Taking the virulent phage T4 as an example, gene expression relies heavily on host transcription and translation machineries. T4 genes can be divided into early, middle and late genes depending on the order through which they are subsequently expressed. Early genes, such as genes encoding

proteins ModA and ModB, usually have strong promoters that compete with host promoters for host transcription machineries. Early gene products further reduce host gene transcription and act

Table 1.1: Summary of different types of cell surface components known to be utilized by phages for adsorption

Host type	Cell surface component	Example	Host	Phage	Reference
Gram-positive bacteria	Peptidoglycan	N-acetyl-muramic acid (MurNAc) of peptidoglycan	*Bacillus thuringiensis*	Bam35	(17)
	Lipoteichoic acids	Negatively charged glycerol phosphate group of Lipoteichoic acids	*Lactobacillus delbrueckii*	LL-H	(18)
	Others	Membrane surface-anchored protein gamma	*Bacillus anthracis*	γ	(19)
Gram-negative bacteria	LPS	Core polysaccharide of LPS	*Pseudomonas aeruginosa*	E79	(20)
	Outer membrane proteins	Protein OmpC	*Salmonella*	Gifsy-1	(21)
	Flagella	Flagellin protein FliC	*Salmonella*	SPN2T	(22)
	Pili	Type IV pili (TFP)	*Pseudomonas aeruginosa*	MPK7	(23)
	Capsule component	Acetyl groups of the Vi exopolysaccharide Capsule	*Salmonella*	Vi I	(24)

to take control of host machineries to express T4 genes. Middle genes are subsequently transcribed 1 minute into infection and mainly encode proteins involved in DNA replication, recombination and metabolism. Middle gene products also include phage-encoded tRNAs and transcription factors that facilitate the switch to late gene transcription (25). Finally, T4 late genes encode components of the phage head, tail, tail fiber and proteins involved in their assembly (26).

In contrast, after DNA injection some phages don't automatically enter the lytic cycle but could, under certain host conditions, enter the lysogenic life cycle. In this cycle, genomic DNA is circularized and integrated into specific sites in the host chromosome or maintained in the host cytoplasm as a circular replicon. In this state, the phage is called a prophage and the host bacterium is called a lysogen. The host lives and reproduces normally, and phage DNA is reproduced and transmitted to daughter cells during cell division. In some cases, prophages can also change the phenotype of the host, for example virulence factors carried on the prophage can turn the host into a virulent bacterium in a process called lysogenic conversion or phage conversion. A well-known example would be the conversion of the non-toxic *Vibrio cholerae* into the highly-toxic cholera toxin-producing version after infection by phage ΦCTX. The lysogenic life cycle continues until, upon exposure to certain conditions such as radiation or mutagenic chemicals, the prophage genome can be excised from the host chromosome and the phage would enter the lytic cycle (27, 28).

1.3.3 Phage assembly and lysis

For virulent phages or temperate phages that have entered the lytic life cycle, assembly of new phage particles occurs after components of the virion are made and genomic DNA is sufficiently replicated. In the case of virulent phages such as T4, assisted by chaperones, the phage assembly process begins with baseplate and later attachment of tails onto the baseplate. The head capsid is assembled separately and packed with genomic DNA before being assembled onto the tails. While packing genomic DNA into head capsids, phages could mistakenly package portions of the host chromosome into their head and inject this host DNA into the cytoplasm of the subsequent host where the DNA from the first host could recombine into the chromosome of the second host. This process is called transduction and is responsible for the dissemination of DNA between bacteria, sometimes the dissemination of virulence or antibiotic resistance genes. In the case of virulent phages, the packaging of host DNA happens randomly, therefore the process is called **generalized**

transduction. In the case of temperate phages, transduction can occur during inefficient prophage excision where the prophage is excised alongside portions of the host DNA. Because prophage integration happens at specific sites in the host chromosome, the host DNA that is excised and packaged into the capsid is also specific. Therefore, this kind of transduction is called **specialized or restricted transduction** (26, 29).

After sufficient numbers of nascent phages are made (e.g. in T4 it could range from 100 to 200 new phage particles), lysis of the host cell occurs. For example, in the case of Gram-negative bacteria, lysis occurs in three main steps where the inner membrane, peptidoglycan layer and outer membrane of the host are targeted by phage-encoded holin, endolysin and spanin complex. Holin accumulates in the cytoplasmic membrane and eventually triggers the formation of holes in the inner membrane. This allows endolysin to escape into the peptidoglycan layer and cause its degradation. Finally, the outer membrane is disrupted by a spanin complex that consists of a small outer membrane lipoprotein and an integral cytoplasmic membrane protein (30). Defects in any of these steps would prevent complete lysis. After lysis of the host cell, phage progenies are released into the environment where they can target new hosts and initiate another cycle of infection.

1.4 Applications of phages and their contribution to molecular biology

Since the discovery of phages a century ago, biologists have studied phages extensively and this has led to our current understanding of many basic biological processes. Phage research has also translated into many applications within the realm of molecular biology itself as well as in biotechnology.

In the early twentieth century, not much was known about genes and their regulation until scientists started using phage as model organisms. For example, in the famous Hershey-Chase experiment conducted in 1952, T2 phages with either their protein coat labeled with ^{35}S or their DNA labeled with ^{32}P were used to infect bacteria. The mixture was agitated to separate phages outside bacteria from bacterial cells, followed by centrifugation to pellet the bacteria. The radioactivity of both the pellet and the supernatant was measured. ^{35}S was only found in the liquid containing free phages that did not infect bacteria or phage parts that did not enter bacterial cells whereas ^{32}P was found within the bacteria cells. This indicated that ^{32}P-labeled DNA entered the cells, not ^{35}S-labeled protein, suggesting that DNA, instead of protein, acts as genetic material (31, 32). Furthermore, phages as model organisms have also contributed to our understanding of gene regulation. For

example, in the 1950s François Jacob's work on λ prophage induction and its genetic circuits led to the identification of DNA-binding repressor and activator proteins and transcriptional anti-terminators (32).

Discoveries in phage biology have also contributed to advancements in molecular cloning and DNA sequencing that revolutionized biological research. For example, in the 1950s the phenomenon of restriction-modification was discovered in bacteria. Bacteria employ restriction endonucleases to cleave foreign DNA, such as invading phage DNA, at specific cut sites as a defense mechanism. In concert, bacteria can also use methylation to modify their own DNA to protect it from restriction. The recognition of specific DNA sequences by restriction enzymes was later used to cleave genes in cloning technology (33, 34). Studies of phages lambda and phage P1 also contributed to the use of cosmids and artificial chromosomes as cloning vectors. Upon the advent of DNA sequencing techniques, phage genomes (e.g. phage ϕX174 in 1977) were the first to be sequenced and contributed to the development of different sequencing methods, such as shot-gun sequencing (4, 32). Studies of the T7 DNA polymerase have also provided a solution for high-fidelity DNA sequencing (35).

More recently, phages have contributed greatly to the development of biotechnological tools. For example, phage display is a method for synthesizing and selecting polypeptides with novel characteristics. In phage display, genes encoding polypeptides are fused with phage coat protein genes and the peptide fusions expressed on the surface of the phage particles. This creates a library of polypeptides that can be screened for desired characteristic such as affinity to target receptors or pathogenic agents (36-38). The application of phage display can be further expanded to enable phages to carry specific genes for targeted gene delivery or vaccine delivery (37). The specificity of phage-host recognition is also used in biotechnology in a technique called phage typing, where phages can be used to detect and identify unknown bacterial strains (39). Finally, phages can be used as therapeutic agents in phage therapy – the clinical use of phages to treat bacterial infection. Phage therapy was first studied in the 1920s by d'Herelle (2, 40), but interest in it waned in the West around the 1940s after the discovery of antibiotics. In the last decade, with growing concerns of antibiotic resistance, interest in phage therapy has resurfaced. Phage therapy boasts several advantages over antibiotics, including the specificity of phages in host recognition that could prevent damage to the healthy gut flora, the localized nature of phage proliferation at the site of

infection, and the ability of phages to reach required target sites such as by crossing the blood-brain barrier (41). However, many obstacles still exist for the clinical application of phage therapy, such as the lack of rigorous clinical research data, the possibility of human immune response against administered phages and, importantly, the threat of anti-phage defense mechanisms employed by bacterial hosts. More research is needed on the basics of phage biology and phage-host interactions.

2. Bacterial defense against phage infection

The infection and killing of bacteria by phages is a very efficient process. This efficiency results in a high turnover of phage particles, which imposes great selection pressure on bacteria to evolve mechanisms that render them insensitive to phage infection. In response, phage mutations that could enable them to infect these insensitive hosts are selected. The perpetual defense and counter-defense dynamic between phage and host can be described as an evolutionary arms race that has helped biologists build an understanding of evolutionary biology. But the possible emergence of phage-insensitive bacteria hosts also poses a threat to potential applications of phage, such as phage therapy. Therefore, in-depth study of anti-phage mechanisms is essential.

2.1 Overview of anti-phage mechanisms

Anti-phage mechanisms target different steps in the phage life cycle, as illustrated in Figure 1.4 (adapted from the review by Seed and colleagues (42)). These include prevention of phage adsorption, hindrance to DNA injection, damage of phage DNA before phage genes can replicate and express and interference with new phage particle assembly.

Successful phage infection begins with phage recognition of bacteria surface receptors. These receptors usually have their own physiological functions but can be utilized by phages for entry. Modulation of the expression of these receptors may have the concurrent effect of limiting receptor accessibility to phages. Decrease in receptor availability can be achieved through phase variation. For example, *Bordetella bronchiseptica* can vary between the Bvg+ phase and the Bvg- phase. Only in the Bvg+ phase can it produce virulence and colonization factors and phage receptors, such as pertactin (42, 43). Bacteria can also limit receptor accessibility by blockage of receptors using self-made molecules or extracellular matrixes. For example, phage T5 adsorption to its host *E. coli* can be prevented by a lipoprotein (Llp) that *E. coli* produces to block the receptor, FhuA. *E. coli* F^+ strains can also produce the outer membrane protein TraT encoded by the F plasmid to modify the conformation of outer membrane protein A (OmpA) that is recognized by many T-even like phages (44, 45). Some bacteria can also produce an extracellular matrix to create a barrier between phages and surface receptors. This strategy can be easily overcome by phages that produce polysaccharide-degrading enzymes such as hydrolases and lyases that destroy the extracellular matrix (46, 47).

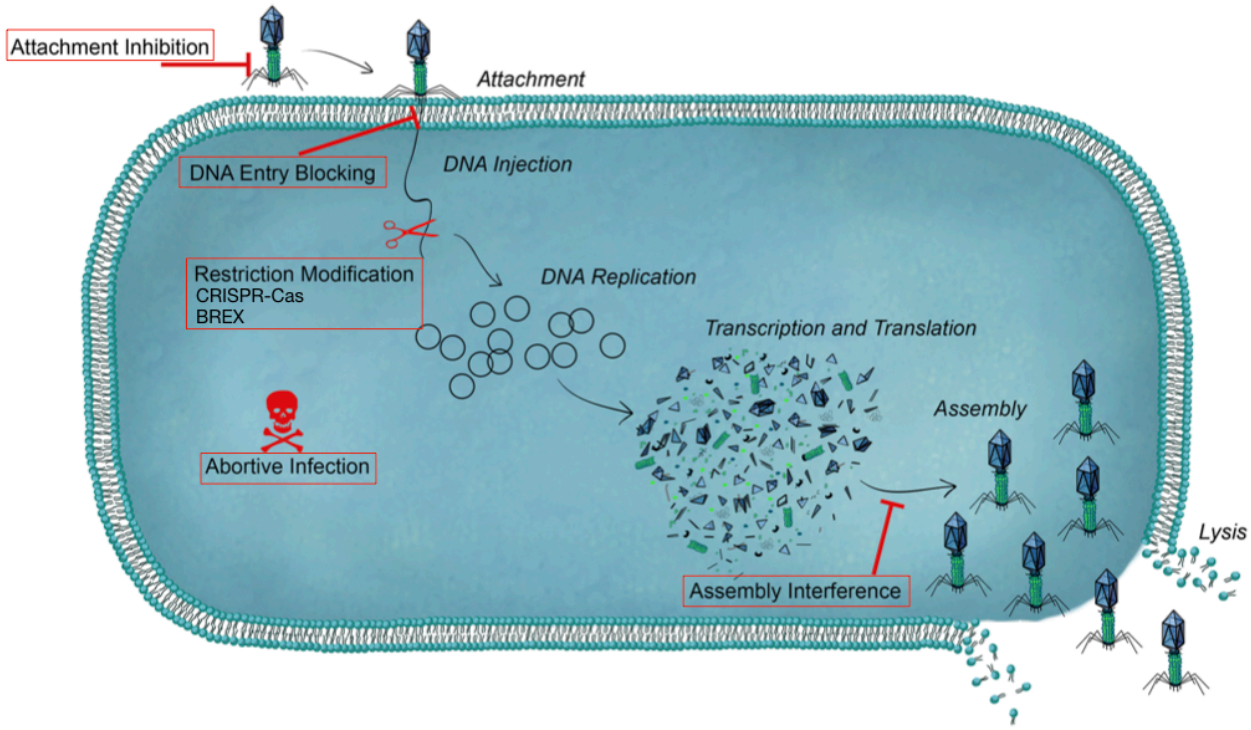


Figure 1.4: Overview of bacterial defense systems against phage. The schematic diagram illustrates anti-phage mechanisms and steps in the phage infection process where they interfere: attachment inhibition interferes with phage adsorption, DNA entry blocking interferes with DNA injection, restriction modification, clustered regularly interspaced short palindromic repeats and CRISPR-associated genes (CIRSPR-Cas), BacteRiophageEXclusion (BREX) all target injected DNA and prevent DNA replication. Abortive infection is considered the "last line of defence" and can target any post-injection steps. Finally, phage particle assembly can also be interfered. The schematic diagram is taken and adapted from the review by Seed and colleagues (42) where CRISPR-Cas and BREX were missing in the original figure. The name of each phage replication step is shown in italics. The name of each anti-phage defense mechanism is highlighted in red boxes.

If the first line of defense is unsuccessful and phages have adsorbed to receptors, phage DNA injection can be prevented by superinfection exclusion (Sie) systems. Interestingly, Sie systems are often encoded by prophages, which can be seen as a part of the host in this context, to avoid secondary infection by another phage. Our understanding of Sie systems is limited but it is believed that these systems encode membrane-associated or membrane-anchored proteins that interact with tape measure proteins of other in-coming phages. Tape measure protein is involved in channel formation for DNA passage through the cell surface, and therefore Sie systems act to block phage DNA from entering the host cytoplasm (42, 48).

Once phage DNA is injected into the host cytoplasm, bacteria can also employ systems such as restriction and modification (R-M), CRISPR-Cas or BREX to cleave viral DNA before phage genes are copied or expressed. R-M systems are usually comprised of a restriction endoribonuclease and a methyltransferase. The restriction endoribonuclease recognizes specific nucleotide bases and cleaves DNA into fragments. Host DNA is methylated on specific residues and thus can be protected from such cleavage. The invading phage DNA is exposed to both the endoribonuclease and methyltransferase at the same time, but due to the higher activity of the endoribonuclease, phage DNA is usually cleaved (49). Some phages have evolved to overcome R-M systems by avoiding endoribonuclease-recognized sequence patterns in their genomes or, such as in the case of T4, by replacing cytosine in its genome with hydroxymethylcytosine to mimic the effect of methylation by a methyltransferase and avoid endoribonuclease recognition (44). In bacteria containing the clustered regularly interspaced short palindromic repeats (CRISPRs) and the CRISPR-associated (Cas) genes, phage DNA is destroyed in a more targeted manner akin to the immune system. The CRISPR locus consists of short DNA repeats separated by short spacer sequences of phage origin and flanked by Cas genes that encode immunization and immunity machineries including nucleases, helicases, polymerases, and polynucleotide-binding proteins. Upon first entry by a phage, a Cas complex recognizes phage DNA and integrates a novel repeat-spacer unit at the leader end of the CRISPR locus. This immunizes the bacteria against subsequent phage infection. If a phage containing identical or similar sequences to any spacer sequence injects its DNA, cognate spacers transcribed into RNAs would help Cas RNA-guided nucleases to locate and target the complementary sequences in the invading phage genome, resulting in phage DNA degradation (50, 51). More recently, a new system called BacteRiophage EXclusion (BREX) was found to also confer bacteria resistance against phages by preventing phage DNA replication and

integration into the host genome while methylating the host cell for self/non-self-distinction. BREX is different from R-M systems because phage DNA is not cleaved or degraded, however the exact mechanism of BREX is still under investigation (52-54).

The later stage of phage particle production and assembly can also be disrupted by bacteria containing staphylococcal pathogenicity islands (SaPIs). SaPIs are 15-17 kb mobile pathogenicity islands in the staphylococcal genome that are also satellite phages. They require helper phages for excision and packaging. To ensure their own progeny production, SaPIs are involved in complicated mechanisms that block late gene transcription and particle production of the helper phage. These mechanisms include interference with full-size helper phage head formation and helper phage DNA packaging (55, 56).

Abortive infection (Abi) is another anti-phage mechanism commonly found in a wide range of bacteria, especially lactococci that are often exposed to phage infection during dairy fermentation (57). Abi differs from other defense mechanism in the following aspects. First, the infected cell does not survive but instead dies or is driven into stasis. This greatly reduces or eliminates the production of phage progeny. Therefore, while most of the anti-phage mechanisms described above protect single bacteria cells, Abi systems arguably protect the entire population of bacteria at the expense of the infected cell. Because of this, abortive infection has been considered a form of "altruistic suicide" and the last line of defense. Finally, instead of targeting a specific step in the phage life cycle, Abi is a collective concept that encompasses a wide range of different defense mechanisms that interfere with earlier post-injection steps during phage proliferation (42). The next section will focus on reviewing our current understanding of Abi mechanisms.

2.2 Abortive infection

Different abortive infection systems have been found to target steps such as DNA replication, gene transcription, mRNA translation, genome packing and phage particle assembly. The effect of Abi is usually manifested by a reduction in Efficiency of Plating (EOP) and decrease in plaque size due to decreased burst size (57). Abi systems have been found in a wide range of bacteria such as *Escherichia coli, Lactococcus lactis* (*L. lactis*)*, Bacillus subtilis, Bacillus licheniformis, Shigella dysenteriae, Streptococcus pyogenes, Vibrio cholerae* and *Pectobacterium atrosepticum*. In particular, more than 20 different kinds of Abi systems have been found on plasmids of lactococci.

Lactococcus lactis converts lactose to lactic acid and is extensively used in fermentation, e.g. the manufacturing of cheese. The fermentation process is often hindered by phage infection of *L. lactis* and therefore anti-phage mechanisms in *L. lactis*, especially Abi systems, have been extensively studied. The majority of Abi systems in *L. lactis* involve a single gene, and protein homology between different systems is usually low. Genetic and biochemical studies in combination with characterization of Abi-resistant *L. lactis* phages have provided insight into the mode of action of a few Abi systems, denoted AbiA-Z. For example, AbiA and AbiK systems, albeit having only 23% protein identity, were found to abort the same family of phages by preventing phage DNA replication, possibly through interference with a phage recombinase. The AbiB system was found to allow normal phage DNA replication and transcription but cause drastic decay of mRNAs. The mechanism of AbiB on mRNA decay is still unclear, the lack of spontaneously resistant phage mutants has limited its study to some extent. The AbiD1 system is active against two highly virulent *L. lactis* phages upon the induction of a certain phage protein, ORF1. AbiD1 interferes with a RuvC-like protein, which is an essential protein for phage propagation because it facilitates the removal of branched DNA generated during replication (58-60). The AbiZ system, found on the same plasmid encoding the AbiA system, could induce early lysis of infected bacteria and reduce phage burst size by 100-fold. The mode of action of AbiZ is believed to involve interaction with holin to induce premature lysis (61).

Many of the characterized Abi systems have been found to be constitutively expressed prior to phage infection. This means that their activity is tightly regulated and inactivated until infection by certain phages activate their function. Unsurprisingly, recent discoveries of new Abi systems have revealed that some of them function as toxin-antitoxin (TA) systems where a toxin “poison” is under the control of an antitoxin “antidote”, such as the discovery of $ToxIN_{Pa}$ in *Pectobacterium atrosepticum*. The discovery that toxin-antitoxin systems could mediate Abi in bacteria have led to re-examination of previously known Abi systems in *L. lactis*, where AbiQ and AbiE were found to function as TA systems as well (62-64).

3 ToxIN$_{Pa}$ – a Type III TA system involved in abortive infection

3.1. Toxin-antitoxin systems

Toxin-antitoxin systems are genetic modules that encode a proteinaceous toxin and an antitoxin that neutralizes the toxicity of the toxin. The antitoxin is more labile than the toxin; its degradation therefore leads to the activation of the toxin. TA systems are encoded in the chromosomes or plasmids of a wide range of bacteria and archaea and are proposed to have versatile functionality important to bacterial growth and survival (65).

TA systems can be categorized into six types based on the nature of the antitoxin and its mode of toxin inhibition. A simplified diagram showing the mechanisms of Type I-VI TA systems are shown in Figure 1.5, taken from the review by Hall and colleagues (66). While toxins in all six types are proteinaceous in nature, antitoxins of different families could be either proteins or RNAs. Type I and Type III antitoxins are RNAs. In the Type I TA system the antitoxin is an antisense RNA that binds to the mRNA encoding the toxin and causes its degradation by RNase pre-translationally. In the Type III TA system, the antitoxin RNA forms pseudoknots and binds directly to the toxin to keep it inactive. Type II, IV, V and VI antitoxins are proteins. In the Type II TA system, the antitoxin binds directly to and inhibits the toxin. In the Type IV TA system, the antitoxin prevents toxin activity by competitively binding to the toxin's target. In the Type V TA system, the antitoxin cleaves the toxin's mRNA directly. In the Type VI TA system, the antitoxin triggers another protein to cleave the toxin (65, 67-69). The mechanisms for the toxicity of toxins also varies by family. Type I toxins are believed to interfere with membrane integrity. Type II toxins are known to interfere with translation by cleaving mRNAs, cleaving initiator tRNAs, binding to ribosomal subunits or phosphorylating elongation factors (67). Type III toxins are endoribonucleases that cleave mRNA at specific sites (70). So far Type I, II, III TAs are relatively well studied while other TA families are represented by one or a few representatives and so require further research.

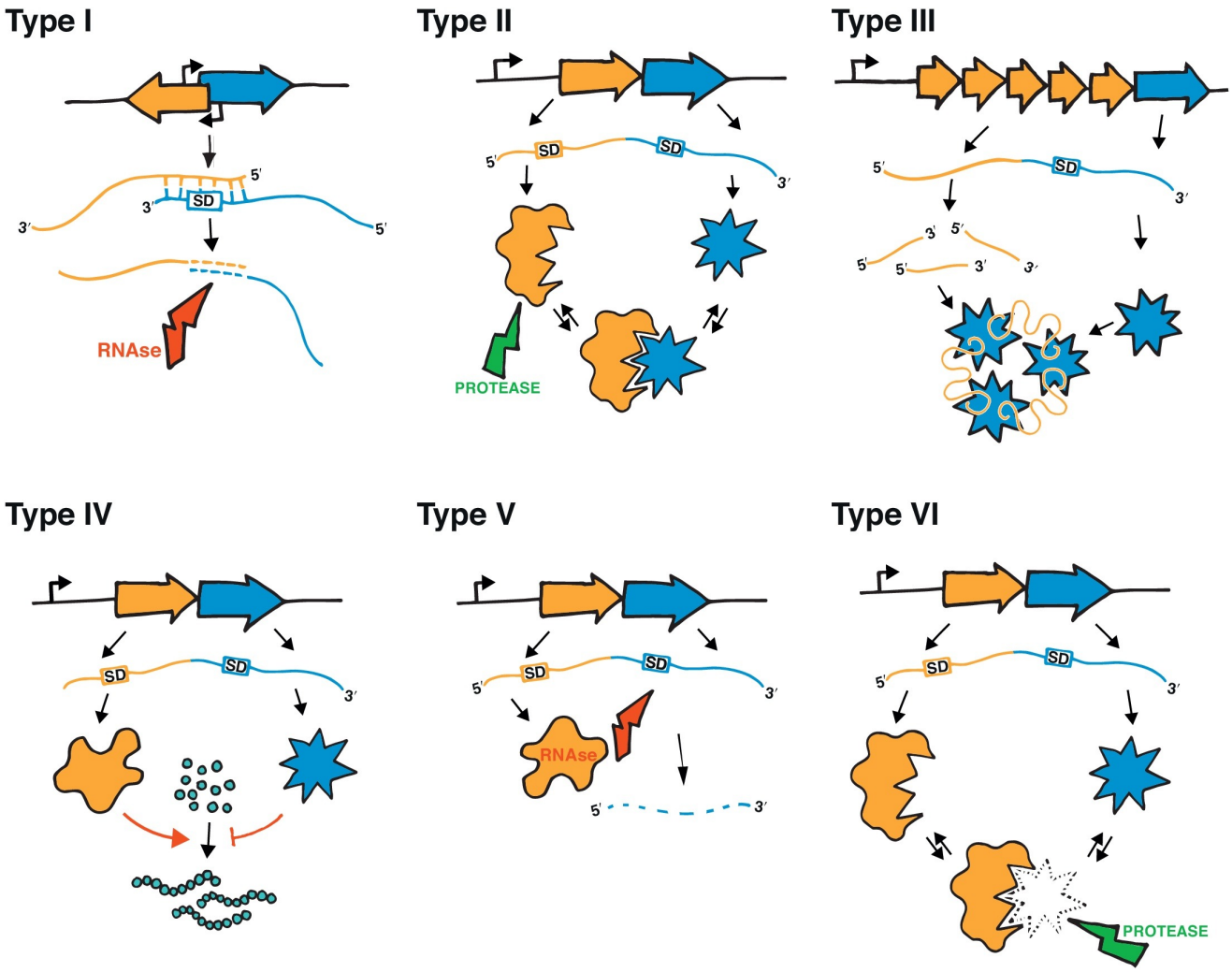


Figure 1.5: Schematic diagram (taken from the review by Hall and colleagues (66)) summarizing the mode of action of Type I-VI TA systems. In the Type I TA system the antitoxin is an antisense RNA that binds to the mRNA encoding the toxin and causes its degradation by RNase pre-translationally. In the Type II TA system, the antitoxin binds directly to and inhibits the toxin. In the Type III TA system, the antitoxin RNA forms pseudoknots and binds directly to the toxin to keep it inactive. In the Type IV TA system, the antitoxin prevents toxin activity by competitively binding to the toxin's target. In the Type V TA system, the antitoxin cleaves the toxin's mRNA directly. In the Type VI TA system, the antitoxin triggers another protein to cleave the toxin (65, 67-69). Blue color represents genes, mRNA transcripts and proteins of toxins. Orange color represents genes, mRNA transcripts and protein/RNA of antitoxins. SD: Shine-Dalgarno ribosome binding site.

The function of TA systems has been studied in depth over the past two decades. When the first TA system was reported in the 1980s, they were found to mediate plasmid maintenance (71). Later work by Gerdes and colleagues found that the *hok/sok* (Type I TA) system also mediated plasmid stability. It was proposed that plasmid maintenance by TA systems occurs through post-segregational killing (72). In this model, if a daughter cell fails to inherit a TA-bearing plasmid, antitoxin from remnant TA products carried over after cell division will be degraded with time, leading to bacteriostasis or cell death caused by the free toxin (72, 73). This "addiction" of plasmid also helps the TA-bearing plasmid outcompete another plasmid of the same incompatibility group if both were acquired through conjugation (69, 74). However, even though the function of TA systems in plasmid maintenance is solidly established, whether or not it functions through post-segregational killing is still debatable as there was no strong evidence showing that physiologically relevant levels of the Hok toxin was lethal (72, 75).

TA systems also promote bacterial survival under unfavorable conditions through the formation of persister cells. For example, the Type II *hipA/hipB* TA module in *E.coli* was the first to be identified as an important determinant for the formation of persister cells (76). Persister cells are a small fraction of cells in a clonal population that enters a reversible dormant state under stress conditions such as exposure to antibiotics, starvation, oxidative stress, pH and extreme temperature. Under antibiotic stress, persister cells maintain minimum metabolic activity. This protects them from elimination since antibiotics mainly target metabolically active cells. Because persistence is a transient phenotypic change instead of permanent genotypic change, persister cells regain activity once antibiotic stress is alleviated. The exact role that TA systems play in inducing persistence is not completely understood. Previous work on persistence-related Type II TA in *E.coli* have suggested that (p)ppGpp controls bacterial persistence by stochastic induction of TA activity (77, 78). However, this model has been overturned recently by retraction of related publications (79).

TA systems are also involved in biofilm formation in both plant and animal pathogens. Biofilms are multispecies consortia embedded in an extracellular matrix that can adhere to biotic and abiotic surfaces (80). In animal pathogens, biofilms are often present in chronic infections, such as due to *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis,* and is crucial for their survival against host immunity. In plant pathogens, biofilm formation is important for survival in the plant

rhizosphere. TA systems are involved in biofilm formation through both the induction of persister cells, which are enriched in biofilm populations compared to planktonic populations (81), and regulation of biofilm formation and dispersal (82). For example the *E.coli* Type II antitoxin MqsA binds to and represses the promoter of the Rpos sigma factor gene *rpos* and the quorum sensing regulator gene, *csgD*, leading to reduced biofilm formation. On the other hand, the cognate toxin MqsR interacts with MqsA to alleviate repression and induce biofilm formation (82, 83).

Furthermore, TA systems are also involved in host stress response. For example, the Type II system *relBE* caused significant reduction in the level of translation after cells experienced amino acid starvation (84), suggesting a role in reducing global translation level during nutritional stress. Work by Wood and colleagues also showed in *E.coli,* MqsA is not only involved in regulation of biofilm formation, but also plays a role in regulating cellular tolerance to antibiotics, oxidative stress and acid stress through repression of *rpoS* (83, 85). In addition, Wood and colleagues have also demonstrated that toxins and antitoxins from disparate systems can be interconverted when they successfully evolved GhoS (an antitoxin from a Type V system) into a novel toxin (ArT) by adding two mutations. ArT dramatically decreased growth in *E.coli*, but mutations introduced to MsqA and ToxI converted these antitoxins from other systems to become antitoxins of ArT (86). These results, taken together with the observed redundancy of TA systems (87), prompted Wood and colleagues to propose that aiding cellular response to a myriad of stresses is the primary function of TA systems (86).

Finally, some TA systems can function as anti-phage mechanisms. For example the Type I TA, *hok/sok* found on *E.coli* plasmid R1, can protect the host from phage T4 infection by forming pores in the membrane (88). Type II TA MazE/F in *E.coli* was found to confer resistance to phage (89). However, it was the discovery of Type III TA complex ToxIN$_{Pa}$ that provided definite evidence that TA systems could provide anti-phage protection through Abi. Subsequently, Type IV TA AbiE was also found to have Abi function (62).

3.2. The discovery and characterization of ToxIN$_{Pa}$

ToxIN$_{Pa}$ was first discovered encoded by pECA1039, a cryptic plasmid of the Gram-negative plant pathogen *Erwinia carotovora* subspecies *atroseptica* 1039 (*Eca* 1039, later reclassified as *Pectobacterium atrosepticum,* hence the subscript in ToxIN$_{Pa}$). Sequencing of the plasmid revealed

an *orf*, denoted *toxN*, that encodes a protein (ToxN) with 31% amino acid identity to AbiQ, an Abi protein in *L.lactis*. This ORF was found to be preceded by another gene annotated *de novo* as *toxI*. Cloning of the region containing *toxN* and *toxI* (*toxIN*) into pBR322 conferred on a related host (*Eca* 1043) resistance against 13 of its 25 phages through Abi. The *toxIN* locus also conferred on *E. coli* DH5α and *Serratia marcescens* Db11 resistance against some of their cognate phages. Individual expression of *toxN* and *toxI* revealed that the product of *toxN* caused bacteriostasis but could be rescued with the expression of the *toxI* locus, therefore indicating its identity as a TA system (64, 90).

Further analysis of the *toxIN* locus revealed its genetic organization (Figure 1.6a). The *toxI* gene and the *toxN* gene are co-transcribed as an operon, with *toxI* sitting upstream of *toxN* and the two genes separated by a rho-independent terminator. The *toxI* gene is composed of 5.5 direct repeats of a 36-nucleotide sequence and transcriptional read-through from *toxI* to *toxN* is reduced to 10% due to a terminator sequence. Experimental data has shown that 2.5 copies of the *toxI* repeat are sufficient, when transcribed from its native promoter, to inhibit ToxN toxicity and function as an anti-phage mechanism (64, 90). The *toxI* gene encodes non-coding RNA that interacts directly with ToxN, albeit being more labile that ToxN. Therefore, the discovery of $ToxIN_{Pa}$ defined a new family of TA systems, designated Type III TA.

Structural studies of $ToxIN_{Pa}$ revealed that three ToxN proteins, each 19.7-kDa, and three ToxI RNA oligomers, each 36-nt, form a heterohexameric, triangular assembly with ToxN proteins sitting at the vertices of the triangle and ToxI oligomers forming the sides. The three ToxI RNAs are pseudo-continuous and each ToxN protein interacts with two ToxI RNAs, the 3’ terminus of one and 5’ terminus of the other. In this complex, each ToxN adopts a globular fold while each ToxI forms a hairpin-type pseudoknot (Figure 1.6b) (70).

Biochemical studies of $ToxIN_{Pa}$ has shown that ToxN is a sequence-specific endoribonuclease and cleaves RNAs at AA↓AU sites (70). The full-length precursor ToxI transcript with 5.5 repeats is also cleaved by ToxN and experimental data suggests that the $ToxIN_{Pa}$ complex is dynamic. Fresh, unprocessed precursor ToxI is constantly cleaved and displaces the mature ToxI already in the $ToxIN_{Pa}$ complex. This cleavage and assembly is likely associated with the inhibition of ToxN toxicity. The dynamic nature of the structure may also give the complex more plasticity in responding to external signals (91, 92).

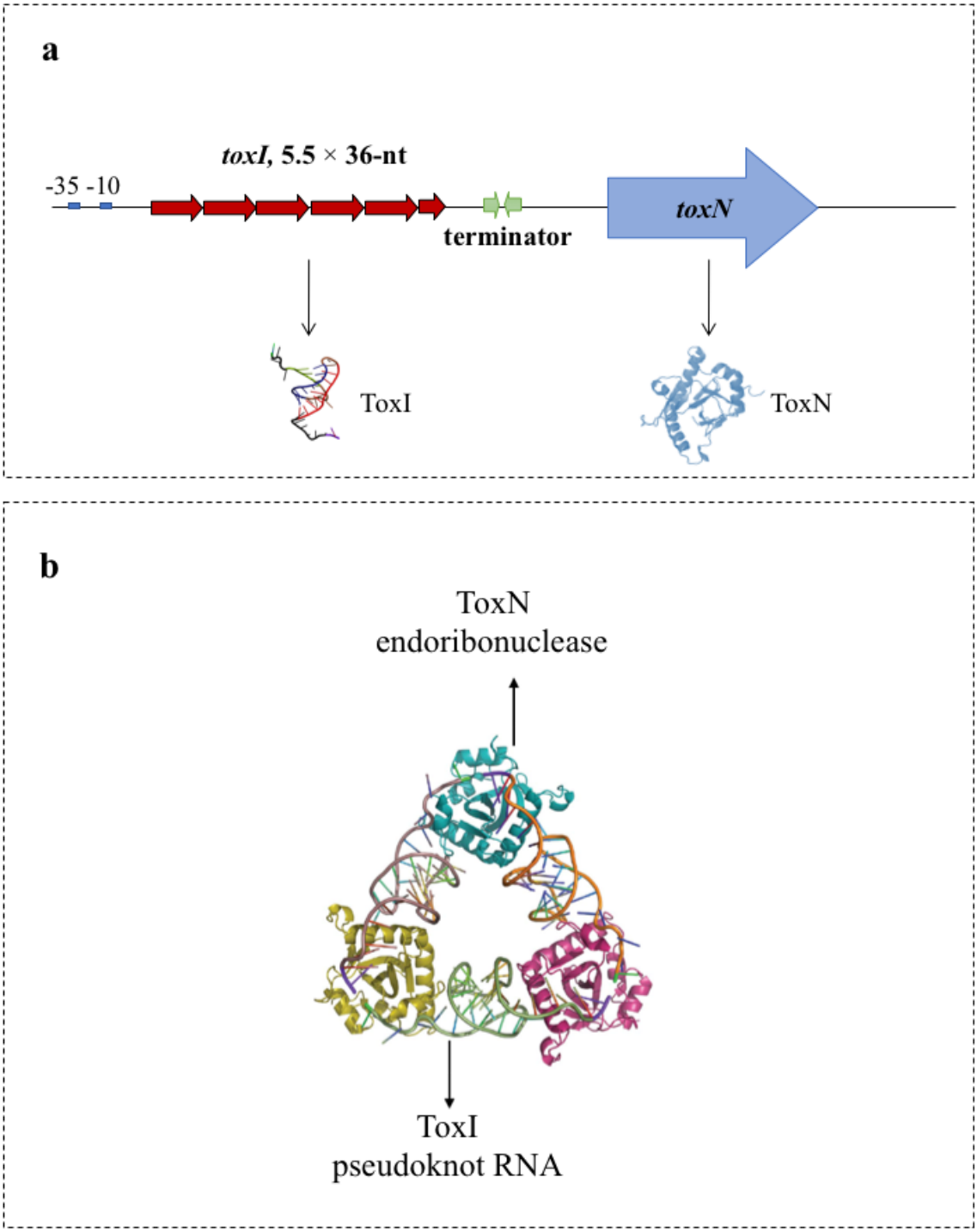


Figure 1.6: Genetic organization of the *toxIN*$_{Pa}$ locus (a) and crystal structure of the ToxIN$_{Pa}$ complex (b). The *toxI* gene and the *toxN* gene are co-transcribed as an operon, with *toxI* sitting upstream of *toxN* and the two genes separated by a rho-independent terminator. The *toxI* gene is composed of 5.5 direct repeats of a 36-nucleotide sequence and transcriptional read-through from

toxI to *toxN* is reduced to 10% due to a terminator sequence. After assembly, the $ToxIN_{Pa}$ complex contains three ToxN proteins, each 19.7-kDa, and three ToxI RNA oligomers, each 36-nt. They form a heterohexameric, triangular assembly with ToxN proteins sitting at the vertices of the triangle and ToxI oligomers forming the sides. The three ToxI RNAs are pseudo-continuous and each ToxN protein interacts with two ToxI RNAs, the 3' terminus of one and 5' terminus of the other. In this complex, each ToxN adopts a globular fold while each ToxI forms a hairpin-type pseudoknot**.** This figure is reproduced from the research article by Blower and colleagues (70).

3.3. Other members of Type III TA

After characterization of the paradigmatic ToxIN$_{Pa}$, attempts were made to find examples of other Type III TA systems using structure-based homology searches with ToxN as the model as well as using the presence of a transcriptional terminator and tandem array of nucleotide repeats as filtering conditions. Hits from the search were further subjected to BLASTp comparison that resulted in separation of the hits into three different families which share low amino acid identity with each other. Based on a representative from each family, the three Type III TA families were named ToxIN (contains the *toxIN$_{Pa}$* locus), CptIN (contains a locus found in *Coprococcus catus* GD/7) and TenpIN (contains a locus found in *Photorhabdus luminescens* subsp. *laumondii* TT01) where the "I" and "N" represent the antitoxin and toxin components respectively (87). Among the total of 125 putative Type III TA systems found in the bioinformatics search, 67 members belong to the ToxIN family, 33 members belong to the CptIN family and 25 members belong to the TenpIN family. In particular, aside from including the TA loci from which they were named, the ToxIN family also includes ToxIN$_{Bt}$ from *Bacillus thuringiensis* and AbiQ from *L. lactis*. The CptIN family also includes CptIN$_{Er}$ from *Eubacterium rectale*. These Type III TA systems are found either on plasmids or chromosomes across a wide range of bacteria (87). Apart from differences in their toxin secondary and tertiary structures, cognate antitoxin genes also differ in length, number of repeats and sequences both within and between families. So far, the structures of ToxIN$_{Bt}$ and CptIN$_{Er}$ have been solved and the structure of AbiQ partially elucidated. Both ToxIN$_{Bt}$ and AbiQ from the ToxIN family adopt a heterohexameric, triangular assembly highly similar to that of ToxIN$_{Pa}$. In contrast, CptIN$_{Er}$ from the CptIN family adopts a heterotetrameric conformation where two toxin monomers are joined by two antitoxin RNAs (91, 93). The structure of TenpIN$_{Pl}$ from *Photorhabdus luminescens* is yet to be reported.

Most of the Type III toxins tested so far, namely toxins of ToxIN$_{Pa}$, ToxIN$_{Bt}$, AbiQ and TenpIN, have been found to act as endoRNases that each recognizes different RNA sequences (91, 94, 95). However, the function of these systems is not uniform. While ToxIN$_{Pa}$, ToxIN$_{Bt}$ and CptIN$_{Er}$ are found to facilitate plasmid addiction, only ToxIN$_{Pa}$ is thus far able to function as an Abi system in the hosts tested (64, 87, 91, 93, 96). AbiQ and TenpIN$_{Pl}$ have also been found to confer phage-resistance on the hosts (87, 97). This has led to the debate of whether phage resistance is the main function of Type III TA or an evolutionary by-product. At the same time, the structural similarity between the ToxIN family members and their different responses to phage infection also presents the mystery as to how Abi-mediating Type III TA modules are activated.

3.4. Phage interactions with Type III TA

Members of Type III TA family, especially ToxIN$_{Pa}$, AbiQ and TenpIN, have been found to facilitate Abi against phages of different families and in different host backgrounds. It is unclear why some phages are aborted by Type III TA while others are not and, for those aborted, no obvious common features have been found between these phage families. Furthermore, it has been observed that phage-aborting Type III TA systems are constitutively expressed, but Abi is only activated upon phage assault, through unknown mechanisms. To better define the selectivity and specificity of Type III TA-mediated Abi, it is important to study the mechanisms of Abi-activation by phages.

Past research on components of the paradigmatic ToxIN$_{Pa}$ have provided hints as to how this complex may be activated by phages. First of all, tracking of ToxN$_{Pa}$ has shown that the toxin level does not increase during phage infection (90, 98), therefore it is unlikely that phages could activate ToxIN$_{Pa}$-mediated Abi by increasing ToxN$_{Pa}$ transcription or translation. Secondly, phage M1 that activates ToxIN$_{Pa}$ was shown to cause a decrease in cellular ToxI$_{Pa}$ by direct or indirect means (98). Thirdly, a mutant ToxN$_{Pa}$ with decreased ToxN$_{Pa}$-ToxI$_{Pa}$ binding was shown to have increased Abi capability (92). Therefore, we could hypothesize several ways in which ToxIN$_{Pa}$-mediated Abi activation may occur: 1) a phage product(s) could perturb *toxIN$_{Pa}$* transcription, the decrease of which could cause the more labile ToxI$_{Pa}$ to be degraded and ToxN$_{Pa}$ freed; 2) a phage product(s) could interact with ToxI$_{Pa}$ to decrease or sequester ToxI$_{Pa}$ and lead to ToxN$_{Pa}$ activation; 3) a phage product(s) could also interact with ToxN$_{Pa}$ to prevent neutralization by ToxI$_{Pa}$ or guide ToxN$_{Pa}$ towards its targets. In addition, phage infection could also trigger the production of a host product(s) or pathway(s) that activates ToxIN$_{Pa}$-mediated Abi in ways similar to the three described above. Due to its dynamic nature, it is likely that ToxIN$_{Pa}$-mediated Abi could be activated in any of these ways with different phages and host background. It is also probable that ToxIN$_{Pa}$-mediated Abi is activated later in the phage infection cycle because normal phage DNA accumulation could still be detected after it takes effect (90).

Direct investigation of ToxIN$_{Pa}$ interaction with phage multiplication steps could only provide limited insight into its activation because it is difficult to distinguish between direct targets of ToxIN$_{Pa}$ and downstream effects. To circumvent this limitation, mutant phages have been used to reveal candidate phage product involved in Abi activation. If a ToxIN$_{Pa}$-sensitive phage is propagated in the presence of ToxIN$_{Pa}$, spontaneous mutants of the phage that are insensitive to

ToxIN$_{Pa}$ could be selected. Comparison of the wild type and mutant phage could then point to product(s) potentially involved in the activation of ToxIN$_{Pa}$-mediated Abi.

Past research on ToxIN$_{Pa}$-insensitive mutant phages have provided insight into the nature of the potential phage product(s) involved. That a phage protein(s) is involved in Abi-activation was shown because phage genes whose mutation could cause ToxIN$_{Pa}$-insensitivity were predicted or proven to be protein-coding. For example ΦM1 of *P.atrosepticum* was sensitive to ToxIN$_{Pa}$ but its mutants could escape ToxIN$_{Pa}$ by acquiring mutations in a gene encoding a conserved, highly toxic protein (M1-23) of unknown function (98). In the similar AbiQ system, several AbiQ-sensitive *L.lactis* phages could overcome Abi also by mutating different protein-coding genes (99). For example, the gene in phage bIL170 encodes a conserved hypothetical protein of unknown function; the early gene *e19* in phage c2 also produces a conserved protein of unknown function; the gene in coliphage 2 produces a protein similar to DNA polymerases; the gene in phage P008 produces an essential phage protein. Furthermore, phage protein products are involved in activation of other types of TA systems (99). However, none of the phage proteins mentioned above share obvious homology with each other, suggesting that the mode of phage interaction with ToxIN systems is diverse.

ToxIN$_{Pa}$-insensitive mutant phages and their mechanisms of escape have also provided insight into phage-host co-evolution. Characterization of the ToxIN$_{Pa}$-sensitive phage ΦTE has shown that mutants could circumvent the ToxIN$_{Pa}$ system by duplicating a DNA sequence highly similar to that of ToxI$_{Pa}$, allowing it to neutralize ToxN$_{Pa}$ and escape. This DNA sequence is present in the ToxIN$_{Pa}$-sensitive wild type phage genome but expanded in the ToxIN$_{Pa}$ -insensitive mutant likely through strand slippage (100). The expansion enabled the production of full-length pseudo-antitoxin to neutralize ToxN$_{Pa}$. This example shows that pressure from host anti-phage mechanisms could select for expansion in the phage's genomic content. Aside from ΦTE, some insight have also been gained from studying ΦM1 of *P.atrosepticum* and its M1-23 protein involved in activating ToxIN$_{Pa}$-mediated Abi. M1-23 was highly toxic, suggesting that activation of ToxIN$_{Pa}$-mediated Abi may be related to toxicity from phage protein(s). Furthermore, mutated M1-23 also co-purified with the host UvrA protein, which is involved in SOS response in bacteria. This may imply that SOS response is related to Abi activation. However, in a *urvA* mutant

background, ΦM1 was still aborted by ToxIN$_{Pa}$, making the mechanism of Abi activation even more elusive (98).

No other ToxIN$_{Pa}$ phages and their escape mechanisms have been completely elucidated. Therefore, it would be interesting to isolate more phages and their corresponding escape mutants to investigate how certain phage product(s) could activate ToxIN$_{Pa}$-mediated Abi, how ToxIN$_{Pa}$ insensitivity could be achieved through mutation, and at what cost.

3.5. ToxIN$_{Pa}$-sensitive phages of *Serratia* sp. ATCC 39006

Serratia sp. ATCC 39006 (*S*39006) is a Gram-negative, rod-shaped bacterium of the genus *Serratia*. It was first isolated in *Salicornia alterniflora* from a salt marsh in Cheesequake, New Jersey (101) and found to be an opportunistic pathogen of a wide range of plant and animal hosts. Recently *S*39006 has attracted growing research interest because of its ability to secrete secondary metabolites such as a carbapenem (102) that could be exploited for future antibiotic development, and prodigiosin, a red pigment and member of the prodiginine class of molecules with antimicrobial, anticancer, and immunosuppressant properties (103). Furthermore, *S*39006 has been found to adopt different modes of mobility beneficial to nutrition acquisition and virulence, such as swimming, swarming and flotation. Because of these characteristics of *S*39006, extensive research has been carried out on regulatory networks in *S*39006. Recently, the genome of *S*39006 has also been fully sequenced and annotated (104).

As a model organism in laboratory where this study was carried out, *S*39006 is devoid of its own Type III TA system and previous isolation of phages infecting *S*39006 resulted in interesting discovery of highly-efficient generalized transducers such as ΦOT8, therefore *S*39006 was chosen as the non-native host to isolate novel phages sensitive to Type III TA systems. *S*39006 was transformed with plasmid-encoded Type III TA systems to test whether these systems have anti-phage activity. Prior to this study, two highly similar phages of *S*39006 had been isolated that could be aborted by ToxIN$_{Pa}$. One of these phages, ΦCHI14, could escape ToxIN$_{Pa}$ via three different routes: mutation of a gene encoding a hypothetical phage protein ORF84; mutation of the phage gene *asiA* encoding the AsiA protein, and deletion of a large genomic region (105). The other phage, ΦX20, isolated a year apart from ΦCHI14, could also escape ToxIN$_{Pa}$ by deleting a large genomic region similar to ΦCHI14 (106). The discovery of ΦCHI14 was the first case in which three different genetic loci could be independently involved the activation of ToxIN$_{Pa}$-

mediated Abi. Furthermore, the high similarity between ΦCHI14 and ΦX20 and their interaction with $ToxIN_{Pa}$, (albeit their independent isolation) hints that common themes exist in *S*39006 phages' response to $ToxIN_{Pa}$. This could be interesting for further exploration to add new members to the diverse repertoire of phage loci involved in $ToxIN_{Pa}$ sensitivity. However, prior to this study, the aforementioned escape loci of *S*39006 phages had not been studied in depth.

4 Aims of this study

Based on the interesting interactions between ToxIN$_{Pa}$ and *S*39006 phages observed in previous studies, one aim of this study was to continue using *S*39006 as the host to isolate new phages from the natural environment. Type III TA systems sub-cloned onto plasmids would be introduced into *S*39006 and phages would be tested for sensitivity to Abi. Type III TA-sensitive phages would be characterized morphologically, genotypically and biologically. This would test the hypothesis that Type III TA-sensitive phages of *S*39006 all belonged to a small clade of highly similar phages, which had already been observed in the case of ΦCHI14 and ΦX20.

After acquiring Type III TA-sensitive *S*39006 phages, another aim of this study was to identify genetic loci involved in their sensitivity by comparing, through whole genome sequencing, wild type phages and their "escape" mutants that are no longer aborted by Type III TA systems. This would test the hypothesis that common themes exist in *S*39006 phages' interaction with Type III TA systems, if similar loci from ΦCHI14 and ΦX20 could be found in newly isolated phages. All of the Abi-activating loci found in *S*39006 phages would be characterized in depth to reveal their biological function and interaction with Type III TA systems.

Previous work had shown that it was easy to isolate Type III TA-sensitive phages for some hosts but difficult for others (Nathalie Goeders, unpublished data; Xinzhe Fang, unpublished data). Therefore, the last aim of this study was to test the hypothesis that Type III TA-mediated Abi is host dependent. The bacterial surface component that *S*39006 phages utilize as receptor would be investigated. Heterologous expression of this receptor in other bacteria would be attempted to expand the host range of these phages and examine their response to Type III TA in new hosts. This would provide insight on the role that the host background plays in Type III TA-mediated Abi.

Chapter 2 - Materials and Methods

1. Media, reagents and solutions

Media used for bacterial growth are listed in Table 2.1. Antibiotics and other supplements used are listed in Table 2.2. Chemical solutions are listed in Table 2.3. Solutions were sterilized by autoclaving at 121°C for 20 minutes or were filter sterilized using 0.22 μm filters.

Table 2.1 Media used in this study

Medium	Components per liter
Luria-broth (LB)	10 g Bacto tryptone 5 g Bacto yeast extract 5 g NaCl
2×LB	20 g Bacto tryptone 10 g Bacto yeast extract 10 g NaCl
Luria-broth agar (LBA)	LB + 15 g agar
0.35% top agar/agarose	LB + 3.5 g agar/agarose
0.75% top agar	LB + 7.5 g agar
Tryptone Swarm Agar (TSA)	10 g Bacto tryptone 5 g NaCl 3 g Bacto agar
Eiken agar	10 g Tryptone 8 g NaCl 2 g Eiken agar 6 g Sodium citrate
0.75% Mueller Hinton top agar	21 g Muller Hinton Broth (Oxoid) + 7.5 g agar 1 L of Muller Hinton Broth gives: 2.0 g Beef extract 17.5 g Casein hydrolysate 1.5 g Starch (pH 7.3 ± 0.1)
1.7% Mueller Hinton agar	1 g Muller Hinton Broth (Oxoid) + 17 g agar

Table 2.2 Antibiotics and supplements used in this study

Chemical	Abbr.	Stock concentration	Working concentration
Antibiotics			
Ampicillin	Ap	100 mg/mL in dH_2O, store at -20°C	100 μg/mL
Chloramphenicol	Cm	50 mg/mL in EtOH, store at -20°C	50 μg/mL
Kanamycin	Kn	50 mg/mL in dH_2O, store at -20°C	50 μg/mL
Tetracycline	Tc	10 mg/mL in EtOH, store at -20°C	10 μg/mL
Supplements			
2,6-Diaminopimelic acid	DAPA	30 mM in dH_2O, stored at 4°C	300 μM
L-arabinose	L-ara	20% in dH_2O, store at 4°C	0.05% or higher
D-Glucose	Glu	20% in dH_2O, store at 4°C	0.2% or higher

Table 2.3 Solutions used in this study

Solution	Components
Phage buffer	10 mM Tris-HCL pH 7.4 10 mM $MgSO_4$ 0.01% Gelatin
Solution A (per L)	9.9 mL 1M $MnCl_2$ 49.5 mL 1M $CaCl_2$ 198 mL 50 mM MES 742.6 mL dH_2O
Solution B (per L)	9.9 mL 1M $MnCl_2$ 49.5 mL 1M $CaCl_2$ 198 mL 50 mM MES 300 mL 50% Glycerol 442.6 mL dH_2O
50× TAE buffer (per L)	242 g Tris base 57.1 mL Glacial acetic acid 100 mL 0.5 M EDTA pH 8.0
Agarose gel (1%)	1 g Agarose in 100 mL TAE buffer 4 µl Ethidium Bromide
Phosphate buffered saline (PBS)	1 PBS tablets (Oxoid) dissolved in 100 mL dH_2O to provide: 0.8 g NaCl 0.02 g KCl 0.115 g Na_2HPO_4 0.02 g KH_2PO_4 pH 7.3
MUG assay reaction mix	In 100 mL PBS: 400 µg/mL Lysozyme from chicken egg white 125 µg/mL 4-Methylumbelliferyl beta-D-galactoside (MUG)

2. Bacterial strains and culture conditions

Bacterial strains used in this study and their growth conditions are summarized in Table 2.4. Strains were grown either in liquid culture shaking at 215 rpm or on LBA plates. Bacterial growth was monitored by measuring optical density at 600 nm (OD_{600}) using a Thermo Scientific Helios Zeta spectrophotometer. Overnight cultures were set by inoculating a single colony into 25 mL universal tubes containing 10 mL LB and the appropriate antibiotic where required. Overnight cultures were grown at the appropriate temperature on a rotating wheel. Stocks of bacterial strains were made by mixing 1 mL of overnight culture with 1 mL of 50% glycerol in a CryoTubeTM (NUNC) and stored in a -80 °C freezer.

Table 2.4 Bacteria strains used in this study

Strain name	Temp.	Relevant characteristics	Source
Escherichia coli strains			
DH5α	37 °C	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ⁻	Gibco
CC118 *λpir*	37 °C	*araD, Δ(ara, leu), ΔlacZ74, phoA20, galK, thi1, rspE, rpoB, argE, recA1, λpir*	(107)
ESS	37 °C	β-lactam sensitive indicator strain	(108)
β2163	37 °C	(F−) RP4-2-Tc::Mu *ΔdapA*::(*ermpir*) Kn^R Em^R	(109)
Serratia strains			
***Serratia* sp. ATCC 39006 LacA (*S*39006 wt)**	30 °C	Laboratory strain, referred to as wild type (wt) in text, Lac-derivative of ATCC *S*39006, carbapenem+, prodigiosin+	(110)
***S*39006-R1**	30 °C	*ompW* mutant of *S*39006 with Tn-KRCPN1 transposon insertion (genotype of Tn-KRCPN1 detailed in Table 2.7)	This study
***S*39006-R4**	30 °C	Spontaneous *ompW* mutant of *S*39006 where the peptide product is truncated to only 47 amino acid remaining at the N-terminal domain	This study
***Serratia marcescens* strain 12**	30 °C	Wild type	(111)

***Serratia plymuthica* A153**	30 °C	Wild type	(112)
Dickeya strains			
***Dickeya solani* MK10**	25 °C	Wild type	(113)
***Dickeya zeae* NCPPB 3532**	25 °C	Wild type	(113)
***Dickeya chrysanthemi* NCPPB 402**	25 °C	Wild type	(113)
***Dickeya dadantii* 3937**	25 °C	Wild type	(114)
***Dickeya dianthicola* NCPPB 453**	25 °C	Wild type	(113)
***Dickeya paradisiaca* NCPPB 2511**	25 °C	Wild type	(113)
***Dickeya* sp. NCPPB 3274**	25 °C	Wild type	(115)
Other strains			
***Pectobacterium carotovorum* SCRI193**	25 °C	Wild type	(116)
***Pectobacterium atrosepticum* SCRI1043**	25 °C	Wild type	(117)
***Citrobacter rodentium* DBS100**	37 °C	Wild type	(118)
***Pseudomonas aeuroginosa* PAO1**	37 °C	Wild type	(119)
***Pantoea agglomerans* 9rz4**	30 °C	Wild type	(120)
***Pectobacterium carotovorum* sp. ATCC 39048**	25 °C	Wild type	(120)

3. Phages, phage techniques and assays

Phages used in this study are summarized in Table 2.5.

Table 2.5 Phage strains used in this study

Phage name	Relevant characteristics	Source
ΦOT8	Transducing phage infecting *S*39006	(120)
ΦCBH8	Wild type	This study
ΦCBH189	Wild type	This study
ΦCHI14	Wild type	This study
ΦX20	Wild type	(106)
ΦCBH8f, i, m, o, p, t, u, x	"Large deletion" mutants of ΦCBH8, details included in Table 3.3	This study
ΦCBH8a, l, r	*asiA* mutants of ΦCBH8, details included in Table 3.3	This study
ΦCHI14w	*asiA* mutant of ΦCHI14 with V13G mutation	(105)

3.1 Isolation and purification of phages from the environment

Phages ΦCBH8 and ΦCBH189 were isolated from treated sewage effluent collected from the sewage treatment plant at Milton, Cambridge. A 5-mL sample of effluent was filter sterilized using 0.22 μm filters (Millipore) and mixed with 5 mL of 2×LB and 500 μl of *S*39006 overnight culture. After overnight incubation, 1 mL of the enriched culture was treated with 50 μl chloroform and subjected to centrifugation at 1,500×g for 5 min. The supernatant was serially diluted, and each diluted lysate mixed with 150 μl of *S*39006 overnight culture and 5 mL of top agarose, and then poured as an overlay onto LBA plates. After overnight incubation, any resulting phage plaques were examined and further purified by picking with a sterile toothpick into 150 μl of phage buffer. The mixture was treated with 20 μl chloroform, centrifuged and the supernatant was serially diluted and plated out again. This process was repeated at least 3 times until all the phage plaques appeared homogeneous.

3.2 Phage titer determination and lysate preparation

To prepare a high titer phage lysate, top agarose samples containing near-confluent phage plaques were scraped off into glass universal tubes containing 2 mL phage buffer and 500 μl chloroform. The mixture was vigorously vortexed and centrifuged at 2,200×g for 20 min at 4 °C. The supernatant was then filter sterilized using 0.22 μm filters (Millipore) and stored at 4 °C. To measure the titer of the phage lysate, the lysate was serially diluted and 10 μl each dilution was mixed with 150 μl of bacterial overnight culture and 5 mL of top agarose, then poured as an overlay onto LBA plates. After overnight incubation at the appropriate temperature, the resulting plaques were counted to determine the plaque forming units per mL (pfu/mL).

3.3 Efficiency of plating (EOP) assay

Phage titer on the test and the control strain was measured separately as described in Section 3.2. The EOP was calculated by [titer on test strain ÷ titer on control strain].

3.4 Phage spot test

An overnight culture of the bacterial host (150 μl) was mixed with 5 mL of top agarose and poured onto LBA plates as a seeded top lawn. The top lawn was left on the bench to set, after which the phage lysate to be tested was serially diluted and 4 μl of each dilution was spotted on the top lawn in order. After the spots had dried, plates were incubated overnight at the appropriate temperature.

If the phage could infect the host, completely clear spots could be observed at high concentration whereas individual plaques could be observed at low concentration. In contrast, if the phage could not infect host, no individual plaques could be observed at lower concentration, but sometimes turbid spots were seen at extremely high concentration (neat samples) due to the effect of "lysis from without" instead of true replicative infection.

3.5 Phage host range assay

Colonies of bacteria to be tested was inoculated in 10 mL cultures with appropriate supplements if needed and grown overnight at the appropriate temperature. The susceptibility of these bacteria to the phage was tested in a standard spot test as described in Section 3.4.

3.6 Phage sensitivity to Type III TA systems

*S*39006 was transformed with a plasmid encoding one of the Type III TA systems (ToxIN$_{Pa}$, ToxIN$_{Bt}$, CptIN$_{Er}$ or TenpIN$_{Pl}$) or one of the frameshifted mutant systems (ToxIN$_{Pa}$-FS, ToxIN$_{Bt}$-FS, CptIN$_{Er}$-FS or TenpIN$_{Pl}$-FS). The infection of these strains by phages was examined by spot tests as described above. If the phage did not form clear spots on the test strain but formed clear spots on the cognate control strain, then the phage was considered potentially sensitive to the Type III TA system. Sensitivity was further confirmed by testing the EOP of the phage infecting the test and control strains. Testing of ΦCBH8 sensitivity to ToxIN$_{Pa}$ in *E.coli* DH5a followed the same procedure, except for addition of 0.05% L-ara and the appropriate antibiotic to the bacterial overnight culture or LBA plates to induce receptor expression and phage adsorption.

3.7 Isolation of ToxIN$_{Pa}$ escape mutants of ΦCBH8

A high titer lysate of wild type ΦCBH8 was serially diluted and each dilution was plated out with an overnight culture of wild type *S*39006 as top lawn. After overnight incubation, the plate with near-confluent phage plaques was chosen to prepare another sample of a high titer ΦCBH8 lysate. This was repeated several times to ensure that spontaneous mutants had enough opportunity to arise in the ΦCBH8 lysate. After several wild type ΦCBH8 lysates had been prepared, each lysate was serially diluted again, and each dilution plated out again, this time with *S*39006 containing ToxIN$_{Pa}$ as the host. After overnight incubation, one plaque from each plate containing visible plaques was picked and prepared as potential escape mutant lysates. To verify that they were true escape mutants, the EOP of these mutants was assessed by plating on *S*39006 containing ToxIN$_{Pa}$

or ToxIN$_{Pa}$-FS. The EOP of true escape mutants should be close to 1.

3.8 Phage transduction

Overnight cultures of the donor strain with appropriate chromosomal or plasmid selection marker were infected in a top lawn with serial dilutions of the phages to be tested for transduction capability. After overnight incubation, phage lysate from the top lawn was harvested as described in Section 3.2. To test the transduction ability of a phage, an overnight culture of the recipient strain was infected with different MOIs of the phage. The phage-host mixture was kept static at room temperature for 20 minutes, followed by 20-minute incubation on the roller wheel at the appropriate temperature. The phage-host mixture was then centrifuged, and the cell pellets were washed twice with LB. After the second wash, cell pellets were resuspended in 200 µl of LB, 100 µl of which was then plated onto LBA plates with appropriate antibiotics. Transduction efficiency was calculated as [the number of transductants that grew on selection plates ÷ the number of cells infected]. If the goal was to use a known transducing phage to transduce chromosomal markers or transfer plasmids, then overnight cultures of the recipient strain were infected with the optimal MOI of the phage to achieve maximum transduction efficiency. After similar washing and plating steps as described above, colonies that arose on selection plates were re-streaked twice. In both cases, recipient-only or phage-only controls were also plated out.

3.9 Adsorption assay

An overnight culture of bacteria was adjusted to OD_{600} = 1 with LB in a 250 mL flask and infected with phages at a MOI of 0.01. The 10 mL phage-bacteria mixture was quickly mixed by vortexing and placed in a water bath at the appropriate temperature with shaking at 150 rpm One hundred microliters samples were taken at different time points and added to 900 µl of chilled LB to reduce re-adsorption. The samples were treated with 50 µl of chloroform and phage titer in the samples was determined. Adsorption curves was drawn by plotting the percentage of free phages in the culture against time. Where needed, an LB-only sample was infected with phages as a negative control.

3.10 One-step burst size assay

Phage burst size assays were modified from the method described by Petty and colleagues (121). In brief, an overnight bacterial culture was used to inoculate fresh LB in a 250 mL conical flask

and incubated at the appropriate temperature to mid-exponential phase. Phage samples were then added at a MOI of 0.1 and the culture incubated with shaking at 150 rpm at the appropriate temperature. One hundred microliter samples were taken at different time points and added to 900 µl of chilled LB to reduce readsorption. Samples were treated with 50 µl of chloroform and phage titer in the samples was determined. The one-step growth curve was drawn by plotting the number of phages per initial infection center against time.

3.11 Phage genome extraction

Phage DNA extraction was modified from the phenol-chloroform protocol described by Sambrook and colleagues (122). In a phase-lock gel (PLG) tube (5' Prime), 450 µl of high titer phage lysate was added with 4.5 µl of 1 mg/mL DNase I and 2.5 µl of 10 mg/mL RNase A and incubated at 37°C for 30 minutes. After incubation, 11.5 µl of 20% SDS and 4.5 µl of 10 mg/mL Proteinase K was added into the tube and the mixture was incubated for 30 minutes. DNA in the mixture was extracted by adding 500 µl of a Phenol-Chloroform-Isoamyl Alcohol (25:24:1) solution and centrifuged at 1500×g for 5 minutes. The supernatant was transferred to a new PLG tube and the previous step repeated. In a new PLG tube, the supernatant was added with 500 µl of Chloroform-Isoamyl Alcohol (24:1) and centrifuged at 1,500×g for 5 minutes. The aqueous phase at the top of the PLG tube was transferred into an Eppendorf tube and incubated with 45 µl sodium acetate (3 mol/L, pH 5.2) and 500 µl of 100% Isopropanol at room temperature for 15 minutes. The mixture was then centrifuged at 12,000×g for 20 minutes. The pelleted DNA was washed at least twice with 70% ethanol and re-suspended in deionized H_2O.

4. Recombinant DNA techniques

All DNA techniques were performed based on standard protocols provided by commercial kit manufacturers, unless stated otherwise. Oligonucleotide primers were obtained from Sigma-Aldrich (now part of Merck KGaA) and are listed in Table 2.6.

4.1. DNA purification and visualization

4.1.1. Bacterial plasmid and genomic DNA extractions

Plasmids used in this study are listed in Table 2.7. Plasmid DNA was purified using the GeneJET Plasmid Miniprep Kit and genomic DNA was purified using the Genomic DNA Purification Kit, all according to the manufacturers' instructions. The concentration and purity of extracted DNA was measured using the NanoDrop™ 1000 Spectrophotometer. All the above kits and instruments were from Thermo Scientific.

4.1.2. Visualization and extraction of DNA from agarose gel

Agarose gels were prepared as described in Table 2.3. A 6× loading buffer (New England Biolabs) was mixed with DNA samples and standard ladders (Bioline) were included for size reference. Gels were run in TAE buffer in an electrophoresis gel tank (Bio-Rad) for 1.5 hours at no higher than 80V. Gels were visualized using a Gene Genius Bio-Imaging System (Syngene, Synoptics Ltd.). To extract DNA from the agarose gel, bands were excised using a scalpel and purified using a GeneJET Gel Extraction Kit (Thermo Scientific) according to the manufacturer's instructions.

Table 2.6 Primers used in this study

Primer	Sequence (5'-3')	Description	Restriction site
pBAD33-F	CTGTTTCTCCATACCCGTT	Forward primer of pBAD33 MCS*	-
pBAD33-R	CTCATCCGCCAAAACAG	Reverse primer of pBAD33 MCS	-
oMAMV1	GGAATTGATCCGGTGGATG	Sequencing primer of pKRCPN1	-
oMAMV2	GCATAAAGCTTGCTCAATCAATCAC	Sequencing primer of pKRCPN1	-

oBH1	GCCTGCTACACCCATCTGG	Primer flanking the largest deleted region in all of ΦCBH8/ ΦCHI14/ ΦX20 mutants combined	-
oBH2	TGGATGCGGTGGTGGATATC	Primer flanking the largest deleted region in all of ΦCBH8/ ΦCHI14/ ΦX20 mutants combined	-
oBH50	GGGGGAGCTCAAGCATACCGAGGTGAACA	Forward cloning primer of LD-ORF8 into pBAD33 MCS	*Sac* I
oBH51	CCCCAAGCTTTTATCTACACTCTTTA	Reverse cloning primer of LD-ORF8 into pBAD33 MCS	*Hin*d III
oBH52	GGGGGAGCTCTACTTGGAGACGTAATGAATAT	Forward cloning primer of LD-ORF9 into pBAD33 MCS	*Sac* I
oBH53	CCCCAAGCTTCTAGCCATTGTTCACCTC	Reverse cloning primer of LD-ORF9 into pBAD33 MCS	*Hin*d III
oBH54	GGGGGAATTCACGGAGTTATATTAAAAC	Forward cloning primer of LD-ORF10 into pBAD33 MCS	*Eco*R I
oBH55	CCCCGGTACCTTATGCTTTCAGAGATT	Reverse cloning primer of LD-ORF10 into pBAD33 MCS	*Kpn* I
oBH42	GGGGGAGCTCACTGAGGAGAACGTTATGAAAAAAAT	Forward cloning primer of LD-ORF11 into pBAD33 MCS	*Sac* I
oBH43	CCCCAAGCTTTTACTTATTTTTAACTCGGC	Reverse cloning primer of LD-ORF11 into pBAD33 MCS	*Hin*d III
oBH148	GGGGGGTACCTTAAAGGAACCAAAATGTTAAAATTCAAA	Forward cloning primer of LD-ORF12 into pBAD33 MCS	*Kpn* I
oBH149	CCCCGTCGACTTATTTACGTGATTTTTCGTAACGAGCAGCGT	Reverse cloning primer of LD-ORF12 into pBAD33 MCS	*Sal* I
oBH158	GGGGGGTACCACGTCTCCGAATTATCTCGTATGTGT	Forward cloning primer of LD-ORF13 into pBAD33 MCS	*Kpn* I
oBH159	CCCCCTGCAGTTATTCTTCGTACCATGGAGTTCTTAAG	Reverse cloning primer of LD-ORF13 into pBAD33 MCS	*Pst* I
oBH160	GGGGGGTACCTAAAGTACATTAATTTTTGAGTGGCCA	Forward cloning primer of LD-ORF14 into pBAD33 MCS	*Kpn* I
oBH161	CCCCCTGCAGTCAATGATATAAAATTAATTGTTCAGTGA	Reverse cloning primer of LD-ORF14 into pBAD33 MCS	*Pst* I
oBH44	GGGGGAGCTCAATGGAGACTATTGATGGATATCGGCTC	Forward cloning primer of LD-ORF15 into pBAD33 MCS	*Sac* I

oBH45	CCCCAAGCTTTCATTGAGCTTGAAGCATTT	Reverse cloning primer of LD-ORF15 into pBAD33 MCS	*Hin*d III
oBH46	GGGGGAGCTCCAAGTAGAGGACATATGAAAAAA	Forward cloning primer of LD-ORF16 into pBAD33 MCS	*Sac* I
oBH47	CCCCAAGCTTTCAATAGTCTCCATTAGCAAGAAAC	Reverse cloning primer of LD-ORF16 into pBAD33 MCS	*Hin*d III
oBH152	CCCCCGTCGACAGCGCTCCCCTCATAAGGGATTGGTCACTGGTTCGAA	Forward cloning primer of LD-ORF17 into pBAD33 MCS	*Sal* I
oBH153	CCCCAAGCTTTCATATGTCCTCTACTTGAGCCATAACAGTGCGTTT	Reverse cloning primer of LD-ORF17 into pBAD33 MCS	*Hin*d III
oBH162	GGGGGGTACCTCCATCGTATAGCGGCTATTATGACTGGCT	Forward cloning primer of LD-ORF18 into pBAD33 MCS	*Kpn* I
oBH163	CCCCCTGCAGTTAGCTACACATAAATTCTGCAACTT	Reverse cloning primer of LD-ORF18 into pBAD33 MCS	*Pst* I
oBH164	GGGGGGTACCTGGTGAATAAATAATTATGAAGACATTGTTA	Forward cloning primer of LD-ORF19 into pBAD33 MCS	*Kpn* I
oBH165	CCCCCTGCAGTTAATTTATATCAAATATGGTGTTG	Reverse cloning primer of LD-ORF19 into pBAD33 MCS	*Pst* I
oBH166	GGGGGGTACCCAAGAGAAATGAAAAATGATGAAACATACTATCGC	Forward cloning primer of LD-ORF20 into pBAD33 MCS	*Kpn* I
oBH167	CCCCCTGCAGTTAGAAACCGAAGCCCACACCTACACCGAATAC	Reverse cloning primer of LD-ORF20 into pBAD33 MCS	*Pst* I
oBH3	CTGTGACTTCGAGCTTAAATCTCC	Forward primer for *asiA* gene in ΦCBH8/ ΦCHI14 genome	-
oBH4	CGCTATATGTCAACAGGCCG	Reverse primer for *asiA* gene in ΦCBH8/ ΦCHI14 genome	-
oBH173	GGGGGAGCTCGCTGTAGGAAAAATAATGTCTAAGA	Forward cloning primer of *asiA* into pBAD33 MCS	*Sac* I
oBH174	CCCCAAGCTTTTACGCGTTATGCATAGCCAG	Reverse cloning primer of *asiA* into pBAD33 MCS	*Hin*d III
oBH182	CCCCGGTACCACCGGTCGAAAGTTCGAATCTTTC	Forward cloning primer of tRNA-Leu into pBAD33 MCS	*Kpn* I
oBH183	GGGGCTGCAGCCACCACGCGTTGGAAGCC	Reverse cloning primer of tRNA-Leu into pBAD33 MCS	*Pst* I
oBH184	CCCCGGTACCTGCAATCTTGGAGGTTCG	Forward cloning primer of tRNA-Arg into pBAD33 MCS	*Kpn* I

oBH185	GGGGCTGCAGGCGGGAGATTTTAAGTCTCCTGT	Reverse cloning primer of tRNA-Arg into pBAD33 MCS	*Pst* I
oBH186	CCCCGGTACCGAGAGGGCTATTTTCAAAGCAA	Forward cloning primer of tRNA-Met into pBAD33 MCS	*Kpn* I
oBH187	GGGGCTGCAGCAATAGATAACAAAATCTTG	Reverse cloning primer of tRNA-Met into pBAD33 MCS	*Pst* I
oBH190	CCCCGGTACCAGAGGTTGGATCGTTACCAA	Forward cloning primer of tRNA-Gly into pBAD33 MCS	*Kpn* I
oBH191	GGGGCTGCAGTTTCGATGTAGTAATAATAACACG	Reverse cloning primer of tRNA-Gly into pBAD33 MCS	*Pst* I
oBH171	CCCCCCGAGCTCAGGAGGAAAGGGGTGGGATGAAAAAGGCGTCTTTATTATTA	Forward cloning primer of $ompW_{S39006}$ into pBAD33 MCS	*Sac* I
oBH172	GGGGGGAAGCTTTCAGAACTTGTAACCTACGGCAAACATAAA	Reverse cloning primer of $ompW_{S39006}$ into pBAD33 MCS	*Hin*d III

*MCS: multiple cloning site

Table 2.7 Plasmids used in this study

Plasmid	Characteristics	Source
pBAD33	Expression vector, repressed by Glu, induced by L-ara, Cm^R	(123)
pBR322	Cloning vector, Ap^R, Tc^R	(124)
pKRCPN1	Vector used for transposon mutagenesis, *tetA, tnp, 'lacZ,* oriR6K, Kn^R	(33)
pTRB30	Kn resistance cassette cloned into pQE-80L NcoI sites	(125)
pTA46	$toxIN_{Pa}$ with native promoter in pBR322, Ap^R	(64)
pTA47	$toxIN_{Pa}$-frameshift with native promoter in pBR322, Ap^R	(64)
pTA114	$toxIN_{Bt}$ with native promoter in pBR322, Ap^R	(64)
pFLS63	$toxIN_{Bt}$-frameshift with native promoter in pBR322, Ap^R	(126)
pFLS116	$cptIN_{Er}$ with native promoter in pBR322, Ap^R	(126)
pTW03	$cptIN_{Er}$-frameshift with native promoter in pBR322, Ap^R	(105)
pFR2	$tenpIN_{Pl}$ with native promoter in pBR322, Ap^R	(95)
pFR2	$tenpIN_{Pl}$-frameshift with native promoter in pBR322, Ap^R	(95)
pBAD33::*orf8*	LD-*orf8* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf9*	LD- *orf9* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf10*	LD- *orf10* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf11*	LD- *orf11* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf12*	LD- *orf12* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf13*	LD- *orf13* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf14*	LD- *orf14* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf15*	LD- *orf15* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf16*	LD- *orf16* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf17*	LD- *orf17* with native RBS in pBAD33, Cm^R	This study

pBAD33::*orf19*	LD- *orf19* with native RBS in pBAD33, Cm^R	This study
pBAD33::*orf20*	LD- *orf20* with native RBS in pBAD33, Cm^R	This study
pBAD33::$tRNA_{Leu}$	tRNA-Leu with native RBS in pBAD33, Cm^R	This study
pBAD33::$tRNA_{Arg}$	tRNA-Arg with native RBS in pBAD33, Cm^R	This study
pBAD33::$tRNA_{Met}$	tRNA-Met with native RBS in pBAD33, Cm^R	This study
pBAD33::$tRNA_{Gly}$	tRNA-Gly with native RBS in pBAD33, Cm^R	This study
pBAD33::*asiA*-wt	*asiA* from wild type ΦCBH8 with native RBS in pBAD33, Cm^R	This study
pBAD33::*asiA*-CBH8a	*asiA* from ΦCBH8a with native RBS in pBAD33, Cm^R	This study
pBAD33::*asiA*-CHI14w	*asiA* from ΦCHI14 with native RBS in pBAD33, Cm^R	This study
pBAD33::*ompW*	Wild type $ompW_{S39006}$ with native RBS in pBAD33, Cm^R	This study
pTA104	$toxIN_{Pa}$ promoter-*lacZ* fusion, pRW50 based, Tc^R	(64)

Notes: Ap^R - ampicillin resistance; Tc^R – tetracycline resistance; Cm^R – chloramphenicol resistance; Kn^R – kanamycin resistance; RBS - Ribosome-binding site

4.1.3. Purification of DNA from enzymatic reactions

After enzymatic reactions such as PCR and restriction digestion, DNA was purified and extracted using the GeneJET PCR Purification Kit (Thermo Scientific) according to the manufacturer's instructions.

4.2. DNA manipulation

4.2.1. Polymerase chain reaction (PCR)

PCR reactions were performed using Phusion® High-Fidelity DNA Polymerase (New England Biolabs). Reactions were carried out in an Applied Biosystems Veriti Thermal Cycler, according to the manufacturer's instructions. Annealing temperatures were adjusted according to specific primer pairs and extension times were calculated by the estimated extension speed of 15-30 seconds per kb. PCR could be carried out using bacterial colonies ("colony PCR") where colonies were picked using a sterile toothpick and added directly to the reaction mixture. An example reaction set-up is shown below:

Reaction mixture:

Component	**20 µl Reaction**	**Final Concentration**
Nuclease-free water	to 20 µl	
5× Phusion HF or GC Buffer	4 µl	1×
10 mM dNTPs	0.4 µl	200 µM
10 µM Forward Primer	1 µl	0.5 µM
10 µM Reverse Primer	1 µl	0.5 µM
Template DNA	variable	< 250 ng
Phusion DNA Polymerase	0.2 µl	1.0 units/50 µl PCR

Program set-up: (Initial denaturation could be extended to 10 min if used for colony PCR)

Step	**Temp**	**Time**
Initial Denaturation	98°C	30 seconds
25-35 Cycles	98°C 45-72°C 72°C	5-10 seconds 10-30 seconds 15-30 seconds per kb
Final Extension	72°C	5-10 minutes
Hold	4-10°C	

4.2.2. Restriction digestion

Digestion of DNA using restriction enzymes (New England Biolabs) was carried out according to the manufacturer's instructions. An example reaction set-up is shown below. Additional enzymes could be used if digesting phage or bacteria genomic DNA.

Component	Volume (µl)
DNA	<1µg
Enzyme	1-2
10× CutSmart Buffer	10
dH_2O	adjust
Total	100

4.2.3. Ligation

Ligation of digested DNA was performed using T4 DNA ligase (New England Biolabs or Thermo Scientific), according to the manufacturers' instructions. The amount of insert and vector DNA to add for optimal ligation was calculated using the New England Biolabs Ligation Calculator, available online.

4.2.4. Sequencing

Sequencing of plasmid and amplicon DNA was performed using the LightRun Sanger Sequencing service at GATC Biotech (now a Eurofins Genomics Company).

4.2.5. Random transposon mutagenesis to select for phage resistant *S*39006

An overnight culture of the donor strain (*E. coli* β2163) was grown at 37°C in LB supplemented with Diaminopimelic acid (DAPA) and appropriate antibiotics. The overnight of the donor strain containing the transposon pKRCPN1 (33) was mixed with the overnight of the recipient strain (*S*39006) in a 1:3 ratio and 30 µl of the mixture was spotted onto LBA plates supplemented with DAPA, but no antibiotic. The mixture was incubated overnight at an appropriate temperature for the recipient strain. The spot containing the mixed culture was then scraped off LBA and re-suspended in sterile LB and spread out onto LBA plates without DAPA but with kanamycin and high titer phage lysate to select for recipient strains that had simultaneously obtained the

transposon and became phage resistant.

4.2.6. Replicon cloning

Replicon cloning was carried as described by Monson and colleagues (33) to identify the insertion position of the transposon Tn-KRCPN1. Genomic DNA of phage-resistant transposon mutants was extracted and digested with the restriction enzyme *Eco*R V that does not cut within the transposon. Digested DNA was purified and subjected to self-ligation. The ligation mixture was used to transform *E.coli* CC118λ*pir* and plated onto LBA containing kanamycin. Replicon DNA from successful transformants was extracted and the location of Tn-KRCPN1 was identified by sequencing using oMAMV1 or oMAMV2 primers.

4.3. Bacterial cell transformation

4.3.1. Preparation of competent cells for heat shock transformation

Overnight cultures of bacterial cells were used to inoculate 25 mL of LB medium containing 15 mM $MgCl_2$ for an OD_{600} of 0.05. Cells were grown at appropriate temperature to mid-exponential phase and then chilled on ice for 60 minutes. Cells were harvested by centrifugation at 4500×g for 10 minutes at 4°C and re-suspended in 10 mL of chilled Solution A. Cell were left on ice for another 20 minutes before being harvested by centrifugation as previously. Finally, cells were re-suspended in 1 mL of chilled Solution B and can be used for transformation or storage at -80 °C.

4.3.2. Preparation of competent cells for electroporation

Overnight cultures of bacterial cells were used to inoculate 25 mL of LB medium for an OD_{600} of 0.05. Cells were grown to mid-exponential phase, chilled on ice for 30 min and then pelleted by centrifugation at 4500×g for 10 minutes at 4°C and washed in 10 mL sterile H_2O pre-chilled on ice. The centrifugation and wash step was repeated, after which the cells were pelleted again and resuspended in chilled 10% glycerol. This wash and centrifugation step was repeated once and cells were finally re-suspended in 1 mL of chilled 10% glycerol, ready for transformation or storage at -80 °C.

4.3.3. Transformation of competent cells by heat shock

Approximately 100 ng of DNA were added to 50 µl of competent cells in an Eppendorf tube and incubated on ice for 20-30 minutes. The bottom 1/2 to 2/3 of the tube was then placed in a heat

block for 2 minutes at 42 °C for heat shock. The tube was then put back on ice for 2 minutes and 900 µl of LB was added. The cells were allowed to recover at appropriate temperature for 1 hour before plating out onto LBA plates with the appropriate antibiotic and incubated overnight.

4.3.4. Transformation of competent cells by electroporation

Fifty microliters of competent cells were added to approximately 50 ng of DNA in an electroporation cuvette on ice. Electroporation was carried out using a Bio-Rad Gene Pulser at 200 Ω, 2.5 kV and 25 µFd. Immediately after pulsing, the cells were supplemented with 1 mL of LB and allowed to recover at the appropriate temperature before plating out onto LBA plates with the appropriate antibiotic, then incubated overnight.

5. Protein expression and toxicity assays

5.1. Protein expression and toxicity assays

Genes to be tested were cloned into the multiple cloning site of the L-ara inducible pBAD33 vector using a standard restriction digestion method described in Section 4.2. Primers used for amplification of the genes are listed in Table 2.6. After successful cloning, recombinant plasmids were introduced into *S*39006 or other hosts by chemical or heat shock transformation. Growth curves were carried out to measure the effect of the cloned gene products in bacteria. In brief, overnight cultures of bacterial cells were used to inoculate 25 mL of LB medium supplemented with the appropriate antibiotic for an OD_{600} of 0.05. Cells were grown in a water bath with shaking at the appropriate temperature to early-exponential phase and gene expression was induced by adding 0.05% or 0.1% L-ara (or higher if needed). Samples were taken at regular intervals and the OD_{600} of samples was measured until the culture reached stationary phase. Where required, the viable count (colony forming units; cfu) of samples was also measured by plating serially diluted cultures on LBA plates supplemented with the appropriate antibiotics. The morphology of cells in the cultures was also monitored, if necessary, by phase-contrast microscopy described in Section 6.2.

5.2. Measurement of ToxIN$_{Pa}$ promoter activity in the presence of AsiA

E.coli DH5α cells transformed with plasmids pTA104 (ToxIN$_{Pa}$ promoter-*lacZ* fusion cloned into pRW50 (64)) or pBAD33::*asiA* were used to inoculate 25 mL of LB medium supplemented with tetracycline and chloramphenicol to an OD_{600} of 0.05. Cells were grown in a water bath with shaking at 37°C until early-exponential phase. AsiA expression was induced by adding 0.05% L-ara and samples were taken every 2 hours, the OD_{600} was recorded and β-galactosidase activity measured in a MUG assay. MUG is the fluorogenic substrate 4-Methylumbelliferyl beta-D-galactoside, which emits fluorescence upon cleavage by β-galactosidase. In brief, 10 µl of culture was transferred to a 96-well plate and frozen immediately at -80°C. Samples were thawed at room temperature and freeze-thawed again to thoroughly lyse the cells. After the final thaw, 100 µl of MUG reaction mix (Table 2.3) was added to each well and samples mixed quickly. Fluorescence was detected in a SpectraMax Gemini XPS fluorescence microplate reader (Molecular Devices) as relative light units per second (RLU/s) using the following settings: excitation 360 nm, emission 450 nm, read intervals 30 seconds/min for 20 min, temperature 37°C. β-galactosidase activity was arbitrarily expressed as RLU/s/mL/OD_{600}.

6. Imaging techniques

6.1. Transmission Electron Microscopy (TEM)

Transmission Electron Microscopy (TEM) images of phages were taken on a Tecnai G2 TEM (Thermo Scientific) at the Multi-Imaging Center, University of Cambridge. Sample grids were prepared by adding 10 µl of high titer phage lysate onto a charged copper grid for 10 min. The grids were then washed with dH_2O twice before being stained with 2% phosphotungstic acid (PTA) neutralized with potassium hydroxide (KOH). The accelerating voltage was 120.0 kV and the direct magnification used to image phages was 25,000×. Measurement of phage dimensions was carried out in ImageJ (NIH).

6.2. Phase contrast microscopy

Phase contrast microscopy images were obtained from 1 µl of bacterial samples from liquid culture taken during a growth curve. Imaging was carried out using an Olympus BX-51 microscope with a 100× oil-immersion lens.

7. OmpW phenotypic assays

7.1. Isolation of spontaneous phage-resistant mutant strains of *S*39006

Isolation of spontaneous phage-resistant mutants was carried out as described previously (127). In brief, an overnight culture of wild type *S*39006 was used to inoculate 5 mL of LB and the culture was grown until OD_{600}=0.5. Following this, the phage lysate was added to achieve a MOI of 1. The phage-host culture was incubated for a further 24 h before 100 µl of the original or diluted culture was plated onto LBA and incubated for 48 h. Five colonies were picked from each sample, re-streaked twice onto fresh LBA plates before being tested for phage resistance in a spot test.

7.2. Secondary metabolite secretion

Prodigiosin

Prodigiosin production was measured as described previously (116). Cells in 1 mL of bacterial culture were pelleted by centrifugation at 13,000 rpm for 10 min. The pellets were re-suspended in 1 mL of acidified ethanol and then vortexed for 5 sec. The cells were pelleted again by centrifugation and 900 µl of supernatant was transferred to a 1 mL cuvette. Prodigiosin was measured by absorbance at 534 nm using a Helios Zeta spectrophotometer (Thermo Scientific). Acidified ethanol was used as blank and relative prodigiosin was represented as $(A_{534}/mL/OD_{600})\times50$.

The Carbapenem

Carbapenem production was measured by comparing the zone of inhibition formed on a top lawn of the carbapenem-sensitive *E. coli* strain ESS. In brief, overnight cultures of the control and test strains of *S*39006 were adjusted to OD_{600}=1 and 3 µl of each strain was spotted on a 0.75% top agar lawn of ESS on the same LBA plate. After overnight incubation at 30°C, the sizes of the halos due to the carbapenem inhibition were compared.

N-butanoyl-L-homoserine lactone (BHL)

BHL production was measured by spotting 3 µl of OD_{600}=1 adjusted overnight cultures of control and test strains of *S*39006 onto an indicator plate. To prepare a BHL indicator plate, 100 µl of the *Serratia* biosensor strain SP19 was made in a 0.75% top agar lawn on an LBA plate. BHL was detected by the production of a red-orange halo by SP19 around the control and test strain colonies (128).

7.3. Swimming and swarming motility assays

Overnight cultures of test and control strains of *S*39006 were adjusted to OD_{600}=0.2 for swimming motility and OD_{600}=1 for swarming motility assays. Three microliters of the cultures were spotted onto test plates. For swimming motility, tryptone swarm agar (TSA) plates were used and incubated at 30°C for 16 hours. For swarming motility, Eiken agar plates were used and incubated at 30°C for 36 hours. Halo sizes were examined to compare the level of motility.

7.4. Stress tolerance assays

Overnight cultures of test and control strains of *S*39006 were adjusted to OD_{600}=1 and serially diluted to 10^{-6}. Ten microliters of each dilution were spotted sequentially on LBA plates supplemented with 2.5% NaCl, 2.5% EtOH or 0.1% SDS (129). After incubation at 30°C for 48 hours, the numbers of colonies were compared for each strain. For susceptibility to H_2O_2, test and control strains of *S*39006 were each made in a seeded top lawn over LBA plates. Fifteen microliters of 3% H_2O_2 was spotted onto a paper disk and placed on the top lawns. After incubation at 30°C for 24 hours, the diameter of inhibition for each strain was measured to compare susceptibility (130).

7.5. Antibiotic susceptibility assays

Antibiotic susceptibility of test or control strains of *S*39006 was tested using the disk-diffusion agar method. In brief, overnight cultures of test or control strains of *S*39006 were made into top lawns with 0.75% Muller Hinton top agar on 1.7% Mueller Hinton agar plates. Commercially available antimicrobial susceptibility disks (Bio-Rad) with standard minimal inhibition concentration (MIC) were placed on the top lawns and plates were incubated at 30°C for 24 hours. The diameter of clearing zone formed by antibiotic inhibition was measured and the inhibition ratio of an antibiotic was calculated as [diameter formed on wild type *S*39006 ÷ diameter formed on mutant *S*39006].

7.6. Virulence assays

Potato infection assay

*S*39006 virulence in potatoes was assessed as described previously (131). Maris Piper potatoes were inoculated with 1×10^6 cfu of wild type or mutant *S*39006 using a clean pipette tip and the point of inoculation was sealed with vacuum grease. The inoculated potatoes were wrapped in wet

tissue paper and cling film before incubation at 30°C. At day 2,3 and 4 of incubation, rotted potato tissue was extracted, weighed and compared.

Caenorhabditis elegans (C. elegans) infection assay

*S*39006 strains to be tested or the control strain *E.coli* OP50 were spotted (10 µl) onto NGM agar in a small Petri dish. After overnight incubation at the appropriate temperature, 50 *C. elegans* DH26 worms synchronized to the L4 stage were picked and added to each bacterial spot. Following overnight incubation at 25°C, the worms were observed under a Stereo Microscope (Leica) and live worms counted. Worms were considered dead if they were no longer responsive to touch. Incubation and counting were repeated daily until all worms had died.

8. Bioinformatics

8.1. Phage whole genome sequencing

The genomes of phages were sequenced using the Illumina MiSeq and HiSeq 2500 platforms at MicrobesNG. Contigs were assembled using Genomic Sequencer *de novo* assembler (Roche) or SPAdes. The wild type sequences had coverage ranging from 57× to 100× of the full genome while the escape phage sequences had coverage ranging from 45× to 150×.

8.2. Phage genome annotation

Phage genome open reading frames (ORFs) were identified using the gene prediction tools GeneMark.hmm (132) and Glimmer (133) with default setting for prokaryotic genomes. ORFs were annotated by searching for homologs using BLASTp (NCBI) searches or i-TASSER (134). The program tRNAScan-SE (135) was used to identify phage tRNA genes and ARAGORN (136) was used to predict host tRNA genes. The GC content of phage genomes was calculated using Artemis (Wellcome Sanger Institute).

8.3. DNA alignment and homology search

Alignment of phage whole genomes was carried out using ACT (Wellcome Sanger Institute) and BLASTn (NCBI). Circular representation of phage genomes was drawn using DNAPlotter (Wellcome Sanger Institute). Pairwise alignment of shorter DNA fragments was carried out using EMBOSS Water at default settings. Search for homologous genes was carried out using BLASTn at default settings.

8.4. Protein alignment and homology search

Alignment of protein sequences was carried out using BLASTp and EMBOSS Needle at default settings. Protein structure and function prediction was also carried out using i-TASSER at default settings. Protein structure visualization was carried out using Swiss-PdbViewer.

8.5. Promoter prediction

Promoter and potential binding sites for regulatory proteins of the *ompW* gene was predicted using the online prokaryotic promoter prediction software BPROM (Softberry) at default settings.

8.6. Codon usage analysis

The codon usage of the phages was analyzed using Artemis by combining all of the ORFs into one entity. The codon usage of bacterial hosts was analyzed by the online tool Codon Usage (http://www.bioinformatics.org/) at default settings.

8.7. Phylogenetic analysis

Phylogenetic analysis of phages was based on similarities of the major capsid protein (MCP). The phylogenetic tree was constructed using Phylogeny.fr (137, 138) at default settings.

Chapter 3 - *S*39006 phages ΦCBH8 and ΦCBH189 activate Type III TA-mediated Abi

1. Introduction

In the natural environment, bacteria are under constant threat of infection from phages. This drives them into an "evolutionary arms race" with phages where both parties involved have come up with mechanisms to ensure continued defense or attack. From the bacterial side (Figure 1.4), mechanisms such as prevention of adsorption, blocking of DNA injection, cleavage of injected DNA via CRISPR-Cas/BREX/restriction modification, and finally, abortive infection (Abi) have been found in evolutionarily distinct bacteria species (44, 54).

Abi, the anti-phage mechanism that causes "altruistic suicide" by bacteria, thereby reducing or eliminating the number of phage progeny produced, occurs via distinct mechanisms in different bacteria. Recently, Abi has been found to be mediated by Type III TA members ToxIN$_{Pa}$ and TenpIN$_{Pl}$ (64, 87, 139). Despite increasing knowledge of Type III TA structures and biochemistry (especially for ToxIN$_{Pa}$) over the years, limited understanding has been gained regarding the interaction of Type III TA systems with phages. Isolation of novel phages from the environment and examination of their interaction with Type III TA systems has shown that only a small proportion of phages could activate Abi mediated by ToxIN$_{Pa}$ and TenpIN$_{Pl}$. Some of these phages also produced "escape mutants" that are insensitive to ToxIN$_{Pa}$, and studies of these mutants have revealed interesting information on how they could activate or escape from ToxIN$_{Pa}$-mediated Abi. For example, two ToxIN$_{Pa}$-sensitive phages, ΦM1 and ΦTE, were isolated against *P. atrosepticum*, the native host in which ToxIN$_{Pa}$ was discovered. ΦM1 encodes a hypothetical protein M1-23 that is believed to be involved in the activation of Abi because different ToxIN$_{Pa}$-insensitive mutants of ΦM1 were found to have mutations in the *orf* encoding the M1-23 protein (98). However, attempts to express and purify M1-23 were unsuccessful due to its high level of toxicity, therefore the exact mechanism of how ΦM1 activates ToxIN$_{Pa}$-mediated Abi through M1-23 is unclear. ΦTE also evolved to produce ToxIN$_{Pa}$-insensitive mutants, which contained several copies of a

DNA sequence highly similar to that of $ToxI_{Pa}$, allowing it to neutralize $ToxN_{Pa}$ and escape (140). However how ΦTE directly activated Abi is also unclear.

Isolation of $ToxIN_{Pa}$-sensitive phages using the native host *P. atrosepticum* provided limited insight into phage interaction with Type III TA systems, therefore other hosts were employed to test whether Type III TA systems are active in non-native bacteria hosts, and to find novel phages that are sensitive. *S*39006 is devoid of its own Type III TA system and previous isolation of phages infecting *S*39006 resulted in interesting discovery of highly-efficient generalized transducers such as ΦOT8, therefore *S*39006 was chosen as the non-native host to study Type III TA systems. Between 2013 and 2014, two M.Phil. students have independently introduced plasmid-encoded Type III TA systems into *S*39006 and isolated two $ToxIN_{Pa}$ and $TenpIN_{Pl}$-sensitive phages. This serves as proof of principle that Type III TA systems can function to abort phages in the non-native host *S*39006.These phages, isolated two years apart, shared extremely high genome sequence identity and shared the same genetic loci that were mutated in their cognate escape mutants (105, 106). However, these escape loci have not been characterized in detail, and no other $ToxIN_{Pa}$-sensitive *S*39006 phages have been isolated since then. Therefore, it is important to try to isolate novel *S*39006 phages from the natural environment.

The study described in this chapter aims at isolating and characterizing novel *S*39006 phages that are sensitive to Type III TA systems. *S*39006 phages will be tested for their EOP when infecting strains containing Type III TA systems and those that are aborted will be chosen for further characterization mainly through whole genome sequencing. Characterization of these Abi-activating phages would enable us to test the hypothesis that these phages are constrained to a small clade of highly similar phages or demonstrate that more distinct phages could be found. Identification of loci altered in "escape mutants" of these phages would also enable us to determine whether similar loci would repeatedly occur, or novel loci could be discovered. This information might point us towards target(s) to focus on for in-depth studies of how these phage loci are involved in Abi-activation.

2. ΦCBH8 and ΦCBH189 activate Abi mediated by $ToxIN_{Pa}$ and $TenpIN_{Pl}$

From October 2014 to May 2015, a total of 175 phages were isolated from the natural environment by enrichment methods using wild type *S*39006 as the host. These phages were tested for their

ability to activate Abi mediated by Type III TA members – ToxIN$_{Pa}$, ToxIN$_{Bt}$, CptIN$_{Er}$ and TenpIN$_{Pl}$. For example, to test phage sensitivity to ToxIN$_{Pa}$, the wild type *S*39006 was transformed with a plasmid (pTA46) encoding ToxIN$_{Pa}$ (64). As control, the wild type *S*39006 was also transformed with a plasmid (pTA47) encoding an inactive version of ToxIN$_{Pa}$ caused by a *toxN* frameshift mutation, denoted ToxIN$_{Pa}$-FS. Phages to be tested were spotted or plated out in full on a top lawn of *S*39006 containing either ToxIN$_{Pa}$ or ToxIN$_{Pa}$-FS. The efficiency of plating (EOP) was calculated as the titer of the phage grown on the ToxIN$_{Pa}$-containing strain divided by the titer of the phage grown on the ToxIN$_{Pa}$-FS-containing strain. Where the EOP was less than 10^{-2}, the phage was considered to activate Abi mediated by ToxIN$_{Pa}$; in other words, "aborted" by ToxIN$_{Pa}$. In a similar way, phage sensitivity to other Type III systems were tested, using wild type and frameshift mutants of the corresponding TA systems.

In the initial assays, two phages, named ΦCBH8 and ΦCBH189, both showed sensitivity to ToxIN$_{Pa}$ and TenpIN$_{Pl}$. As shown in Figure 3.1a, when ΦCBH8 and ΦCBH189 were serially diluted and spotted on ToxIN$_{Pa}$-FS or TenpIN$_{Pl}$-FS, the phages were able to form clear spots, indicating complete lysis. This is expected as the frameshifted versions of these toxins have lost their activity. In contrast, when ΦCBH8 and ΦCBH189 were serially diluted and spotted on ToxIN$_{Pa}$ or TenpIN$_{Pl}$, no clear spots were observed. The turbid spots observed at low dilution factor (high phage titer) is a result of "lysis from without" caused by high concentration of phages outside bacterial cells instead of lysis caused by productive phage lytic infection. This result indicates that ΦCBH8 and ΦCBH189 are unable to propagate in the presence of ToxIN$_{Pa}$ or TenpIN$_{Pl}$ because they have been aborted. In addition to spot tests, plating serially diluted ΦCBH8 and ΦCBH189 in full on an entire top lawn and calculation of their EOP also confirmed that these phages were aborted by ToxIN$_{Pa}$ and TenpIN$_{Pl}$. The EOP of ΦCBH8 and ΦCBH189 is shown in Table 3.1. Consistent with previously characterized ΦCHI14 and ΦX20, ΦCBH8 was aborted by ToxIN$_{Pa}$ to a similar degree and produced spontaneous "escape mutants" that became insensitive to ToxIN$_{Pa}$. ΦCBH8 was also strongly aborted by TenpIN$_{Pl}$ but did not produce any detectable spontaneous escape mutants. On the contrary, ΦCBH189 was strongly aborted by both ToxIN$_{Pa}$ and TenpIN$_{Pl}$ but failed to produce any escape mutants with either system Figure 3.1b. No phages tested have shown sensitivity to ToxIN$_{Bt}$ or CptIN$_{Er}$.

Table 3.1 EOP[a] of *S*39006 phages with ToxIN$_{Pa}$ and TenpIN$_{Pl}$

Phage name	EOP with ToxIN$_{Pa}$	EOP with TenpIN$_{Pl}$
ΦCBH8	8.0×10^{-6}	$< 1.0 \times 10^{-9}$
ΦCBH189	$< 1.0 \times 10^{-9}$	$< 1.0 \times 10^{-9}$
ΦCHI14	2.0×10^{-7}	$< 1.0 \times 10^{-9}$
ΦX20	2.8×10^{-8}	$< 1.0 \times 10^{-9}$

[a]EOP = phage titer on test strain / phage titer on control strain.

Test strain: *S*39006 expressing ToxIN$_{Pa}$ or TenpIN$_{Pl}$.

Control strain: *S*39006 expressing ToxIN$_{Pa}$-FS or TenpIN$_{Pl}$ –FS.

EOP is represented as $< 1.0 \times 10^{-9}$ when the phage is strongly aborted and fails to produce any observable escape mutants.

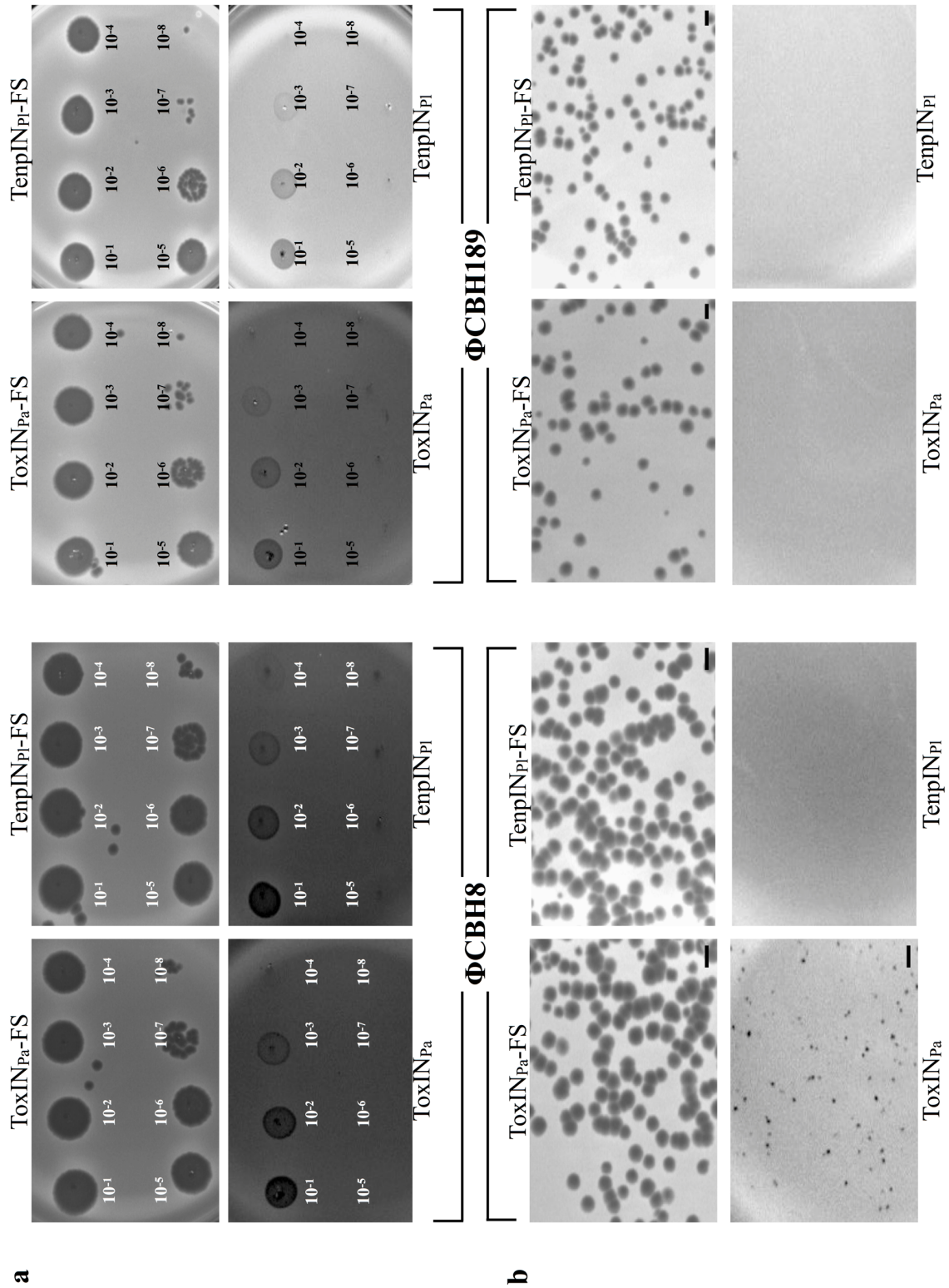


Figure 3.1 (figure legend provided overleaf): ΦCBH8 and ΦCBH189 infecting *S*39006 containing ToxIN$_{Pa}$ or TenpIN$_{Pl}$.

Figure 3.1: ΦCBH8 and ΦCBH189 infecting *S*39006 containing ToxIN$_{Pa}$ or TenpIN$_{Pl}$.
(a) Spot tests of ΦCBH8 and ΦCBH189: when ΦCBH8 and ΦCBH189 were serially diluted and spotted on ToxIN$_{Pa}$-FS and TenpIN$_{Pl}$-FS (top row), the phages were able to form clear spots, indicating complete lysis. When ΦCBH8 and ΦCBH189 were serially diluted and spotted on ToxIN$_{Pa}$ or TenpIN$_{Pl}$ (bottom row), no clear spots were observed. The turbid spots observed at low dilution factor (high phage titer) are due to "lysis from without" instead of productive phage lytic infection. (b) Plaque formation of ΦCBH8 and ΦCBH189: when ΦCBH8 and ΦCBH189 were serially diluted and plated as entire top lawns, both could form clear, round plaques on ToxIN$_{Pa}$-FS and TenpIN$_{Pl}$-FS when the phage lysates were sufficiently diluted (10^{-6}). However only ΦCBH8 was able to produce plaques of "spontaneous escape" mutants in the presence of ToxIN$_{Pa}$ when high titer non-diluted phage lysate was used. No mutants were observed when high titer ΦCBH8 was plated on TenpIN$_{Pl}$, or when high titer ΦCBH189 was plated on either ToxIN$_{Pa}$ or TenpIN$_{Pl}$. Scale bars represent 2 mm. For both ΦCBH8 and ΦCBH189, the non-diluted phage lysate used in (a) and (b) had a titer of ~10^{10} pfu/mL.

3 ΦCBH8 and ΦCBH189 define a new genus of T4-like phages

The fact that ΦCBH8 and ΦCBH189 are both sensitive to $ToxIN_{Pa}$ and $TenpIN_{Pl}$, but demonstrated differences in their ability to circumvent the systems made them interesting subjects for further characterization and comparison. For both ΦCBH8 and ΦCBH189, their morphology, genomic content and similarity with other phages were investigated.

3.1 ΦCBH8 and ΦCBH189 are phages of the *Myoviridae* Family

In order to obtain high resolution Transmission Electron Microscopy (TEM) images of ΦCBH8 and ΦCBH189, high titer phages were stained with PTA (see Section 6.1, Chapter 2). As shown in Figure 3.2, both phages have polyhedron heads, tails and tail fibers. Furthermore, the tails of both phages have also been captured in a contracted state. This provides clear evidence that ΦCBH8 and ΦCBH189 belong to the *Myoviridae* family of the order *Caudovirales*, as seen with ΦCHI14 and ΦX20.

Detailed measurement of ΦCBH8 showed that ΦCBH8 has a head width of 84 nm, head length of 108 nm, extended tail length of 113 nm, extended tail width of 18 nm, while ΦCBH189 has a head width of 78 nm, head length of 110 nm, extended tail length of 110 nm, extended tail width of 24 nm. These measurements indicate that ΦCBH8 and ΦCBH189 are similar in size to ΦCHI14 and ΦX20, also phage T4. The large size of the phage particles can be an indication of large genome size, but since phage morphology could only provide limited information, ΦCBH8 and ΦCBH189 were then subjected to whole genome sequencing.

3.2 Whole genome sequencing of ΦCBH8 and ΦCBH189

Results from whole genome sequencing showed that ΦCBH8 has a genome size of 171,151 bp, with a GC content of 38.74%, much lower than the host GC content of 49.24%. ΦCBH8 has 275 predicted open reading frames (ORFs) that are predicted to encode proteins and 17 tRNA genes. A full list of all ORFs, tRNAs and their predicted function can be found in Appendix I. Of the 275 protein-coding ORFs, 18% are predicted to encode structural proteins or proteins involved in phage assembly while 56% encode hypothetical proteins with no known function. Compared to ΦCBH8, ΦCBH189 has a slightly larger genome size of 172,248 b.p., with a GC content of 38.72%. ΦCBH189 has 276 ORFs that are predicted to encode proteins and encodes for 15 tRNAs. A full

list of all ORFs and their predicted function can be found in Appendix II. Of the 276 protein-coding ORFs, 16% are predicted to encode structural proteins or proteins involved in phage

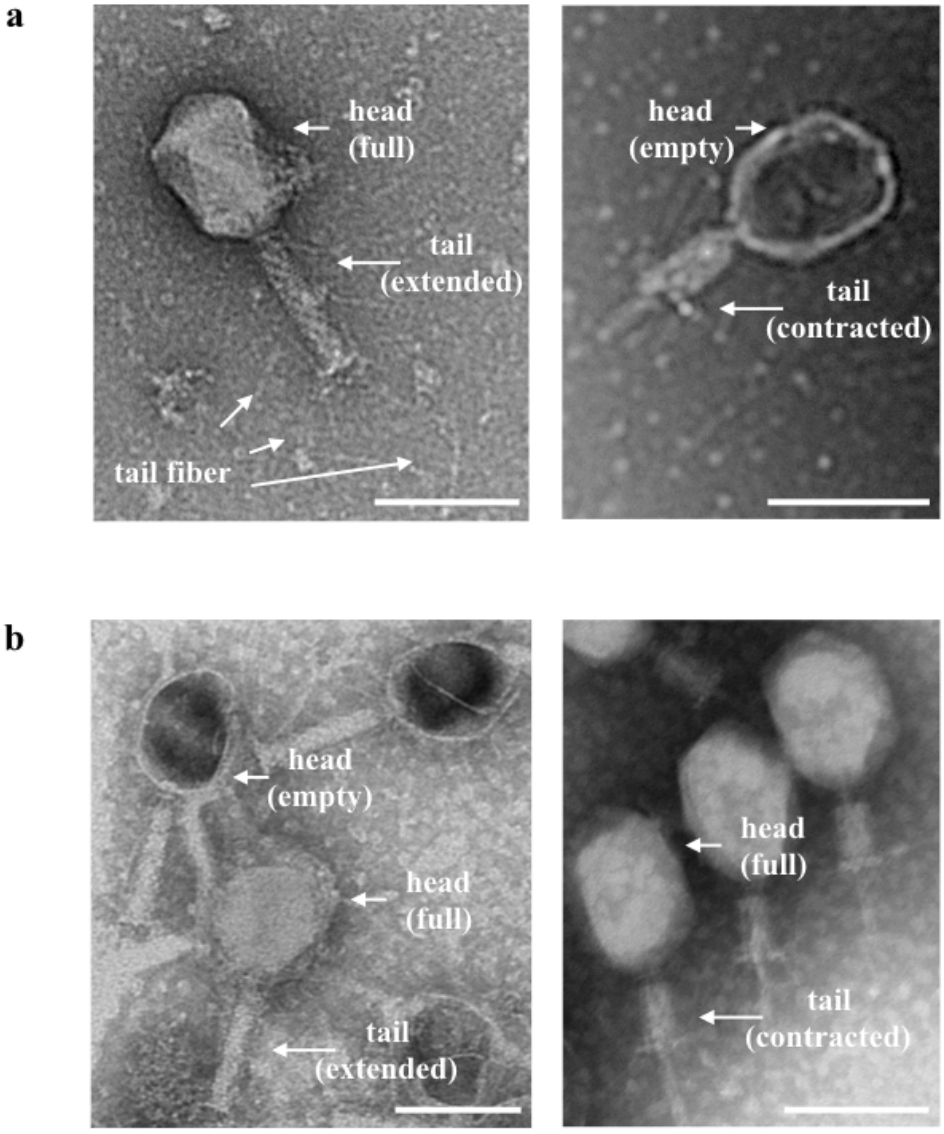


Figure 3.2: (a) TEM images of ΦCBH8 with extended tail (left) and contracted tail (right). (b)TEM images of ΦCBH189 with extended tail (left) and contracted tail (right). Recognizable phage structures are pointed out with arrows. Scale bars represent 100 nm.

assembly while 54% encode hypothetical proteins with no known function. For both phages, the remaining ORFs with known function encode for proteins predicted to be involved in DNA replication, metabolism and other regulatory aspects of phage propagation.

The genomic composition and arrangement of ΦCBH8 and ΦCBH189 were further analyzed by plotting ORFs in a circular representation. As shown in Figure 3.3, ΦCBH8 and ΦCBH189 have ORFs located on both strands, with the forward strand (arbitrarily assigned) carrying all of the tRNA genes, the majority of hypothetical proteins, proteins involved in DNA/RNA replication/metabolism, other regulatory proteins, and a small fraction of structural proteins. These ORFs are interspersed on the forward strand except for the region positioned between approximately 50 kb to 75 kb where ORFs encoding hypothetical proteins are more concentrated. The reverse strand encodes the majority of structural proteins clustered together in three different regions of the genome. The above genomic organization is shared in ΦCBH8 and ΦCBH189.

3.3 ΦCBH8 and ΦCBH189 represent a new species and genus of T4-like phages

To investigate the relationship between ΦCBH8, ΦCBH189 and other phages whose genomes have been sequenced, a nucleotide BLAST search was conducted to look for phages with high whole-genome identity with ΦCBH8 and ΦCBH189. Unsurprisingly, ΦCHI14 and ΦX20 were top hits and an *E.coli* phage EcS1 came in third. As expected, ΦCBH8 and ΦCBH189 also share a high degree of nucleotide identity with each other. Some of the most significant hits from the BLASTn search are shown in Table 3.2. Based on percentage identity and coverage, it seems that ΦCBH8, ΦCBH189, ΦCHI14 and ΦX20 all belong to a new group of phages that share >95% sequence identity with each other over >95% of the genome. This meets the ICTV's requirement for new species classification (141). Furthermore, *Escherichia* virus EcS1 shares >75% sequence identity over >85% coverage of the genome with these *S*39006 phages. On the contrary, other phages in Table 3.2 share a moderately high level of identity with *S*39006 phages but only in limited regions of the genome (overall identity <50% after adjusting for coverage). These phages include the unclassified *Edwardsiella* phages and some classified phages belonging to different genera of the *Tevenvirinae* subfamily, such as phage T4. All of the phage mentioned above belong to the T4-like superfamily due to their genomic similarity of varying degrees to phage T4. This suggests that ΦCBH8, ΦCBH189, ΦCHI14 and ΦX20 belong to the T4-like phage superfamily but could, at the same time, define a new genus that also includes EcS1, based on the ICTV's

criteria that viruses within a genus must share >50% overall nucleotide similarity and be distinct from viruses of other genera (141).

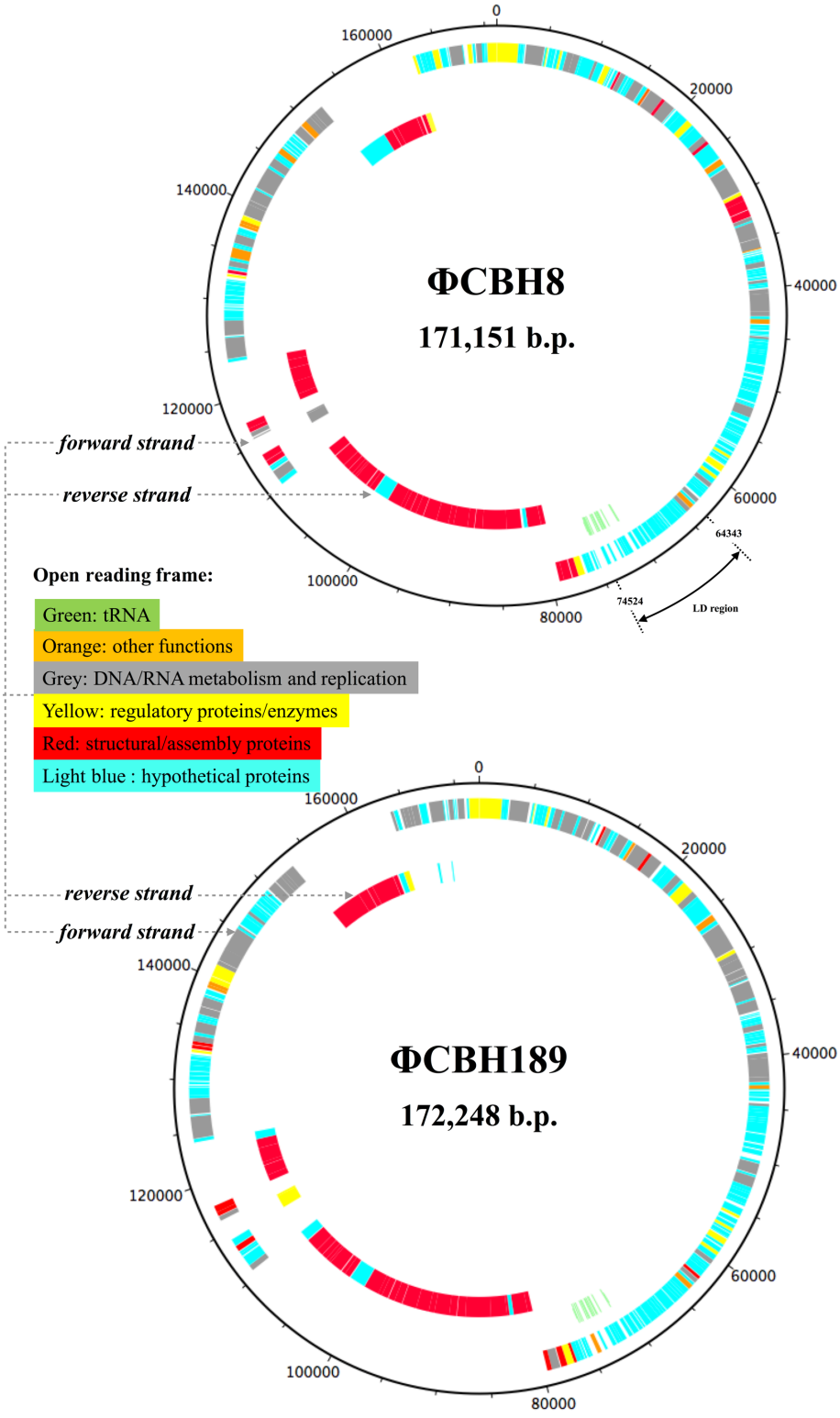


Figure 3.3 (figure legend provided overleaf): Circular representation of ΦCBH8 and ΦCBH189 genomes.

Figure 3.3: Circular representation of ΦCBH8 and ΦCBH189 genomes. The outer ring represents ORFs on the forward strand while the inner ring represents ORFs on the reverse strand. ORFs are color-coded based on their categorized function. tRNA genes are located on the reverse strand but are represented here between the inner and outer strand for easier visualization. The figures are plotted using DNAPlotter (Wellcome Sanger Institute). The "LD region" labeled in the circular map of ΦCBH8 is mentioned later in Section 2.1, Chapter 4.

Table 3.2: Whole genome alignment of ΦCBH8, ΦCBH189 and T4 with related phages using BLASTn

Phage name	**Subfamily : Genus**	**Percentage identity with ΦCBH8 (%) and coverage (%)**		**Percentage identity with ΦCBH189 (%) and coverage (%)**		**Percentage identity with T4 (%) and coverage (%)**	
		Identity	Coverage	Identity	Coverage	Identity	Coverage
ΦCBH8	Unassigned	100	100	98	97	74	49
ΦCBH189	Unassigned	98	97	100	100	74	49
ΦCHI14	Unassigned	99	100	98	97	74	49
ΦX20	Unassigned	99	97	97	97	73	49
Escherichia virus EcS1	Unassigned	78	86	78	85	71	50
Edwardsiella virus PEi20	Unassigned	72	67	71	67	72	47
Edwardsiella virus PEi26	Unassigned	72	67	71	67	72	47
Escherichia virus CC31	*Tevenvirinae : Cc31virus*	70	58	70	58	72	52
Enterobacter virus PG7	*Tevenvirinae : Cc31virus*	73	58	73	57	72	54
Escherichia virus JS10	*Tevenvirinae : Js98virus*	75	53	75	53	76	56
Escherichia virus VR5	*Tevenvirinae : Js98virus*	75	54	75	54	76	58
Escherichia virus JS98	*Tevenvirinae : Js98virus*	75	54	75	55	76	58
Escherichia virus IME08	*Tevenvirinae : Js98virus*	68	54	68	54	76	60
Escherichia virus T4	*Tevenvirinae : T4virus*	74	49	74	49	100	100

To further substantiate the relationship between these *S*39006 phages and other related phages, phylogenetic analysis was conducted using the amino acid sequences of major capsid protein (MCP) of *S*39006 phages and other phages with homologous MCP to them according to BLASTp (142). As seen in Figure 3.4, all the *S*39006 phages belong to the same clade, consistent with BLASTn comparison of their whole genomes. Phage EcS1 is phylogenetically closest to *S*39006 phages and might share a common ancestor. On the contrary, *S*39006 phages are more distant to other phages of the T4-like superfamily.

As *gp23* encoding MCP is a signature protein in T4-family phages, the phylogenetic analysis provides further proof these *S*39006 phages represent a new species and even possibly a new genus of T4-family phages (141). Recently, the name of the new genus has been proposed to the ICTV

(pending approval) as *Winklervirus* and falls under the subfamily of *Tevenvirinae* (Andrew Kropinski, personal communication).

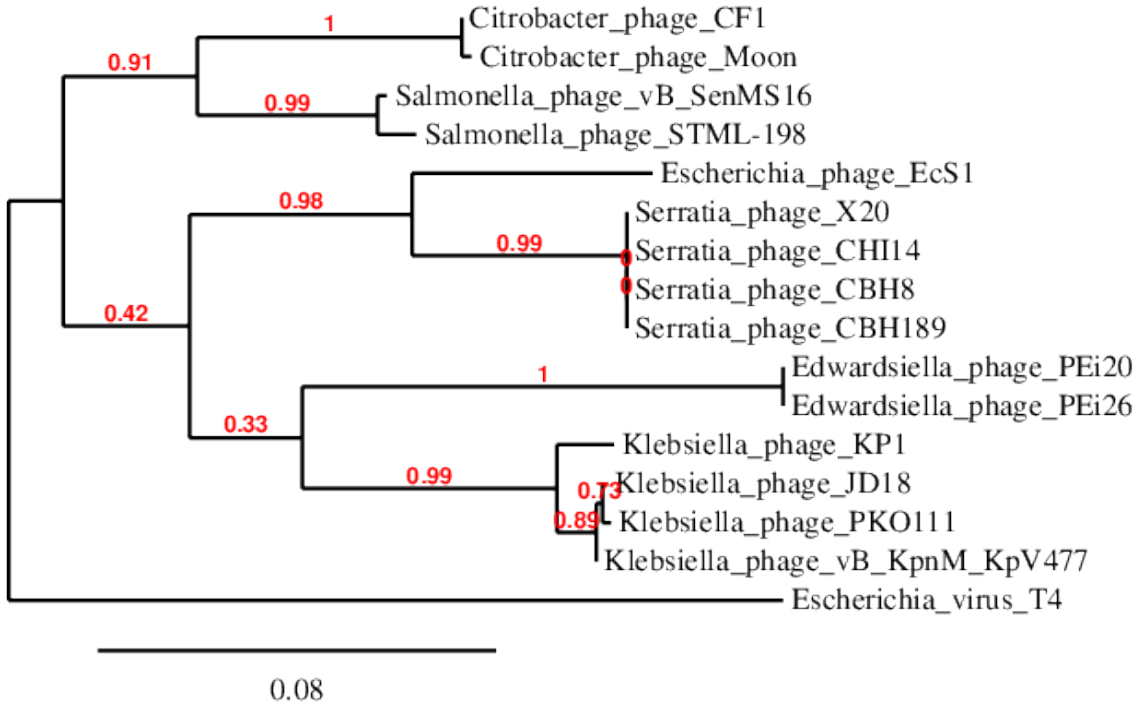


Figure 3.4: The phylogenetic relationship of *S*39006 phages with other phages based on the amino acid sequences of major capsid protein. The evolutionary distance between any two leaves in the tree is proportional to the sums of the branch lengths connecting them. The scale bar (0.08) shows the approximate percentage difference in amino acid sequences associated with that particular length. The tree is unrooted and bootstrap values are shown in red.

4. ΦCBH8 and ΦCBH189 have narrow host range

The host ranges of ΦCBH8 and ΦCBH189 were investigated in order to determine whether these phages infect bacteria other than *S*39006 and potentially to explore their activation of Type III TA systems in another host background. ΦCBH8 and ΦCBH189 were used in spot tests on top lawns containing different bacteria. The bacterial hosts to be tested were chosen depending on their availability in the lab collection as well as their Gram-negative nature. Different species of *Serratia*, *Citrobacter, Dickeya, Escherichia, Pectobacterium* and *Pseudomonas* were tested for their susceptibility to ΦCBH8 and ΦCBH189. However, no bacteria could be infected by these *S*39006 phages, similar to the host range assays carried out previously with ΦCHI14 and ΦX20. The results indicate that this new genus of *S*39006 phages exhibits high host specificity; thus far they only infect the host from which they were isolated.

5. ΦCBH8 could escape ToxIN$_{Pa}$ via distinct routes

Infection assays shown in Figure 3.1 indicated that while ΦCBH8 infection was aborted by ToxIN$_{Pa}$, a tiny proportion of its progeny were able to escape ToxIN$_{Pa}$. Based on an EOP of ~10^{-6}, roughly 1 in a million phages in the ΦCBH8 phage stock were ToxIN$_{Pa}$-insensitive mutants, observed in Figure 3.1b. These mutants would have arisen during the propagation of wild type ΦCBH8 and presumably should contain mutated versions of *orf*(s) involved in the activation of ToxIN$_{Pa}$-mediate Abi. Therefore, attempts were made to isolate these escape mutants and identify their mutated gene(s).

To ensure that each mutant isolated was independent and that the mutant bank created has the greatest diversity, wild type ΦCBH8 was serially diluted and plated out on top lawn of *S*39006 containing ToxIN$_{Pa}$-FS. As shown in the schematic diagram in Figure 3.5, individual plaques of wild type ΦCBH8 progenies were individually picked and "bulked up" again on *S*39006 (ToxIN$_{Pa}$-FS) to give ample opportunities for mutants to arise. These wild type ΦCBH8 progenies were named ΦCBH8wt1, ΦCBH8wt2, ΦCBH8wt3,…,ΦCBH8wtn. Each wild type ΦCBH8 progeny phage stock was then serially diluted and plated out separately on *S*39006 (ToxIN$_{Pa}$) to select for escape mutants. One mutant from each wild type stock was selected, purified and named ΦCBH8a, ΦCBH8b, ΦCBH8c,…, ΦCBH8n. These mutants were plated out on *S*39006 (ToxIN$_{Pa}$) and *S*39006 (ToxIN$_{Pa}$-FS) again to calculate their EOP, all of which were restored to ~1, indicating that they were phages that became true escape mutants.

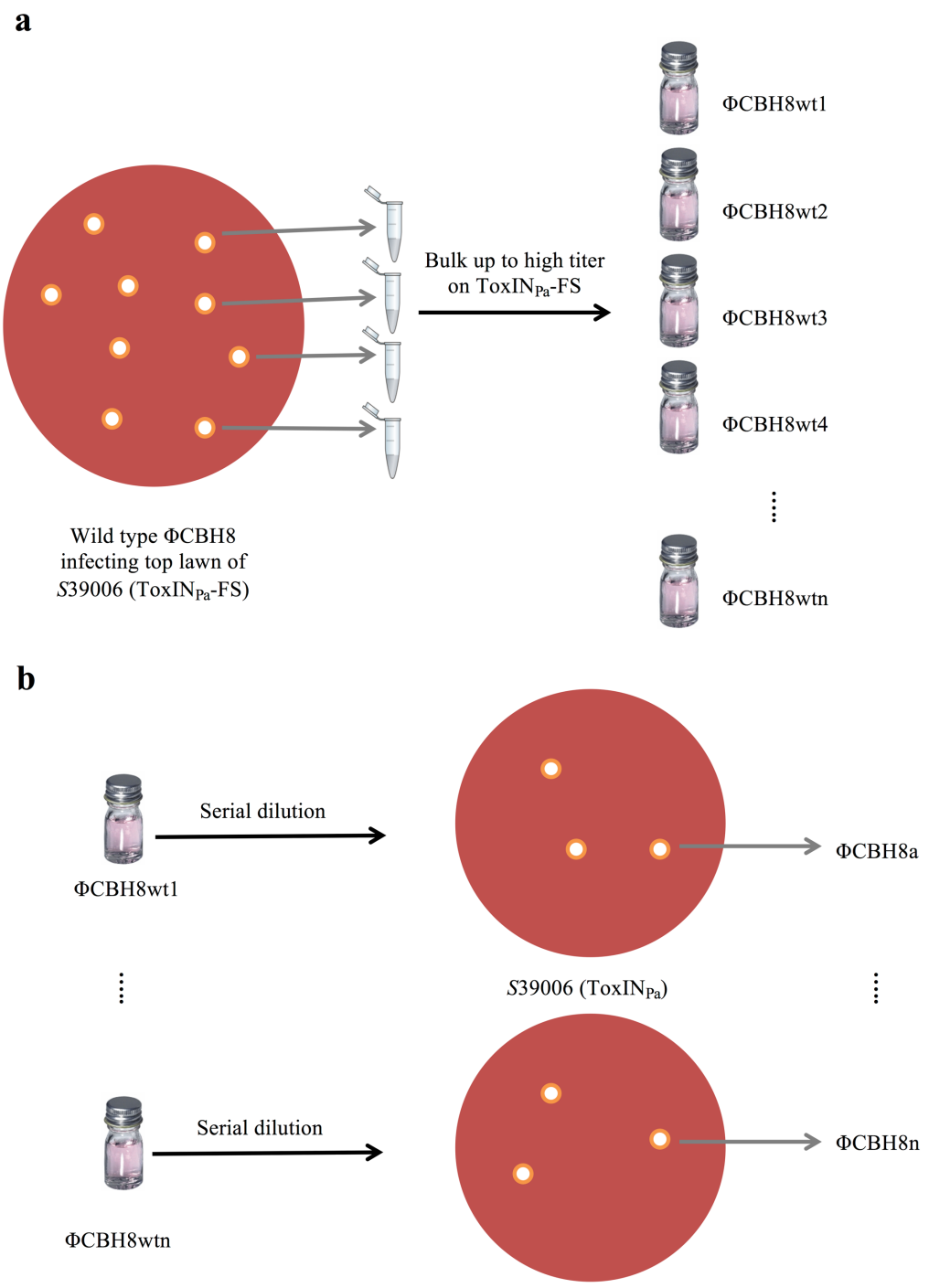


Figure 3.5: Schematic diagram showing the isolation and purification of ΦCBH8 mutants. (a) Plaques of wild type ΦCBH8 progenies were individually picked and bulked up again on *S*39006 (ToxIN$_{Pa}$-FS). These wild type ΦCBH8 progenies were named ΦCBH8wt1, ΦCBH8wt2, ΦCBH8wt3,…,ΦCBH8wtn. (b) Each wild type ΦCBH8 progeny phage stock was then serially diluted and plated out separately on *S*39006 (ToxIN$_{Pa}$) to select for mutants. One mutant from each wild type stock was selected, purified and named ΦCBH8a, ΦCBH8b, ΦCBH8c,…,ΦCBH8n.

A total of 11 independent ΦCBH8 mutants were subjected to whole genome sequencing. Mutant genomes were compared with the wild type genome to identify the mutated gene(s) involved in escaping $ToxIN_{Pa}$ (Table 3.3). Interestingly, the mutations did not all map to the same location in the corresponding genomes. Instead, 8 of the mutants harbored a large deletion (size ranging from 6.5 kb to 10 kb) in their genomes while 3 mutants carried a nonsense mutation in the *asiA* gene. Similar mutations have been found previously in ΦCHI14 and ΦX20 mutants, which is unsurprising given their high DNA identity with each other. However, the details of each mutated locus differ between mutants of ΦCBH8, ΦCHI14 and ΦX20, as shown in Table 3.3. What is worth noting is that the large deletion regions in some phage mutants, albeit having different start and end positions, all overlapped with each other, as shown in detail in Figure 4.1, Chapter 4.

Taken together, the repeated appearance of similar escape loci in these phages suggest variations on a common theme for *S*39006 phages that activate $ToxIN_{Pa}$-mediate Abi. More detailed characterization of these escape loci will be the focus of Chapter 4 and Chapter 5.

Table 3.3. Summary of mutations in $ToxIN_{Pa}$-escape mutants of ΦCBH8, ΦCHI14 and ΦX20.

Escape phage	Mutation type	Gene(s) affected	Effect of mutation
ΦCBH8 mutants			
ΦCBH8f	large deletion (10040 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64343 bp – 74382 bp)
ΦCBH8i	large deletion (7802 bp)	15 *orf*s, 11 tRNAs	elimination of affected genes (66568 bp – 74369 bp)
ΦCBH8m	large deletion (8575 bp)	17 *orf*s, 6 tRNAs	elimination of affected genes (64427 bp – 73001 bp)
ΦCBH8o	large deletion (6521 bp)	13 *orf*s, 11 tRNAs	elimination of affected genes (67248 bp – 73768 bp)
ΦCBH8p	large deletion (8368 bp)	16 *orf*s, 10 tRNAs	elimination of affected genes

			(65513 bp – 73880 bp)
ΦCBH8t	large deletion (7328 bp)	15 *orf*s, 9 tRNAs	elimination of affected genes (65310 bp – 73637 bp)
ΦCBH8u	large deletion (8158 bp)	17 *orf*s, 4 tRNAs	elimination of affected genes (64543 bp – 72700 bp)
ΦCBH8x	large deletion (7731 bp)	14 *orf*s, 10 tRNAs	elimination of affected genes (65959 bp – 73689 bp)
ΦCBH8a	nonsense mutation	*asiA*	E71 → stop codon
ΦCBH8l	nonsense mutation	*asiA*	E72 → stop codon
ΦCBH8r	nonsense mutation	*asiA*	E72 → stop codon
ΦCHI14 mutants			
ΦCHI14a	large deletion (7647bp)	14 *orf*s, 12 tRNAs	elimination of affected genes (66821 bp – 74468 bp)
ΦCHI14b	nonsense mutation	*asiA*	E71 → stop codon
ΦCHI14c	large deletion (10094bp)	19 *orf*s, 13 tRNAs	elimination of affected genes (64429 bp – 74524 bp)
ΦCHI14e	missense mutation	*orf*84	E66D
ΦCHI14f	deletion (10bp)	*asiA*	extends protein by 17 residues
ΦX20 mutants			
ΦX20b	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20d	large deletion (9479 bp)	19 *orf*s, 8 tRNAs	elimination of affected genes (64093 bp – 73572 bp)
ΦX20f	large deletion (9473 bp)	19 *orf*s, 8 tRNAs	elimination of affected genes (64110 bp – 73583 bp)
ΦX20g	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20h	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20j	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)

ΦX20k	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20l	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20m	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20n	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)
ΦX20o	large deletion (9533 bp)	19 *orf*s, 11 tRNAs	elimination of affected genes (64664 bp – 74196 bp)

6. ΦCBH189 cannot escape ToxIN$_{Pa}$

Despite having extremely high overall genome identity with ΦCBH8, ΦCHI14 and ΦX20, ΦCBH189 did not produce any detectable escape mutants (Figure 3.1b, Table 3.1). This is surprising given that at least 3 different escape loci were found among the other *S*39006 phages. Therefore, bioinformatic analysis was carried out to verify whether the ToxIN$_{Pa}$ escape loci found in ΦCBH8 are present in ΦCBH189.

Alignment of *asiA* sequences showed that ΦCBH189 only differs in the 144th and 176th nucleotide of *asiA* compared to ΦCBH8 and ΦCHI14. However, these mutations do not alter the cognate amino acids encoded, and these nucleotides are not the same as those mutated in the escape mutants. Therefore, it is intriguing why the *asiA* mutation has not been observed in ΦCBH189. Comparison with the deleted region in ΦCBH8o (denoted "CBH8o-LD", the smallest deleted region among all mutants) showed that ΦCBH189 also contains this region, with 82% nucleotide identity. Therefore, it is also intriguing why the large deletion mutation failed to arise in ΦCBH189. Further analysis showed that each *orf* and tRNA within CBH8o-LD shares high level of similarity with the *orf* or tRNA in ΦCBH189, with two exceptions. In one case, a predicted hypothetical protein is only present in the cognate region in ΦCBH189 but not in ΦCBH8. In another case, a hypothetical protein in CBH8o-LD is missing in ΦCBH189. It is possible that this *orf* (named *orf*12) was involved in the activation of Abi by ΦCBH8, therefore deletion of the region encompassing *orf*12 produced escape mutants while the absence of *orf*12 explained the lack of escape mutants of ΦCBH189.

Aside from having differences with escape loci found in ΦCBH8, it could also be that ΦCBH189 lacks certain proteins required for the genesis of *asiA* mutation or large deletions. A detailed comparison was made between ΦCBH8 and ΦCBH189 to identify protein-coding regions with known function that differ between the two phages. Table 3.4 summarizes ORFs with significant differences between ΦCBH8 and ΦCBH189 (ORFs with minor alterations between ΦCBH8 and ΦCBH189, and ORFs with significant differences in ΦCBH8 and ΦCBH189 but without known function are omitted). ΦCBH8 and ΦCBH189 possess different types of homing endonucleases, and the Endonuclease IV they possess have different amino acid sequence in the C-terminal domain. However whether these proteins are involved in facilitating *asiA* and large deletion mutation in ΦCBH8 mutants is not clear.

Table 3.4: ORFs of known function that differ in ΦCBH8 and ΦCBH189

Gene name	Source	Difference
segA homing endonuclease	Present **only in ΦCBH8**	Entire *orf* missing in ΦCBH189
tRNA-Lys (TTT)	Present **only in ΦCBH8**	Entire *orf* missing in ΦCBH189
tRNA-Glu (TTC)	Present **only in ΦCBH8**	Entire *orf* missing in ΦCBH189
segB homing endonuclease	Present **only in ΦCBH189**	Entire *orf* missing in ΦCBH8
segD homing endonuclease	Present **only in ΦCBH8**	Entire *orf* missing in ΦCBH189
Long tail fiber, distal subunit	Present in **both ΦCBH8 and ΦCBH189**	82% amino acid identity, similar length protein but CTD[a] did not align
Endonuclease IV	Present in **both ΦCBH8 and ΦCBH189**	82% amino acid identity, similar length protein but CTD[a] did not align

[a]*CTD: C-terminal domain*

[b]*NTD: N-terminal domain*

Since ΦCBH189 did not produce mutants escaping $ToxIN_{Pa}$, the next 2 chapters will focus on characterization of escape loci found in ΦCBH8 and exploration of their relationship with $ToxIN_{Pa}$.

7. Discussion

Over the course of 3 years, > 300 phages of *S*39006 were isolated and tested for their sensitivity to Type III TA systems. However, so far only 1% of the phages screened were aborted by $ToxIN_{Pa}$ and $TenpIN_{Pl}$. This rarity may be unsurprising, since wild type *S*39006 does not contain any Type III TA systems in its genome, and in this study, all of the Type III TA systems were introduced artificially into *S*39006 on a plasmid. Despite the artificial introduction, it was still possible to find rare phages that were sensitive to $ToxIN_{Pa}$ and $TenpIN_{Pl}$. It could be a coincidence that these phages carried ORF(s) that made them susceptible to $ToxIN_{Pa}$ and $TenpIN_{Pl}$. It is also possible that

these phages infect other unknown host(s) that naturally possess a similar Type III TA system. However, screening of a large number of bacteria strains available in the lab yielded no positive result and more research needs to be done to determine whether these ToxIN$_{Pa}$/TenpIN$_{Pl}$-sensitive *S*39006 phages infect other hosts.

None of the >300 phages screened were aborted by ToxIN$_{Bt}$ or CptIN$_{Er}$. This resonates with earlier studies and supports the hypothesis that ToxIN$_{Bt}$ and CptIN$_{Er}$ might not abort phages, and that the original function of Type III TA system could be something other than Abi, such as plasmid addiction (96).

Two of the *S*39006 phages isolated in this study were able to activate Abi mediated by both ToxIN$_{Pa}$ and TenpIN$_{Pl}$. Characterization of the two phages alongside previously isolated ΦCHI14 and ΦX20 showed that while they share some level of similarity with phages of the T4-like superfamily, they are more similar with each other than with these T4-like phages. Whole genome alignment and phylogenetic studies using major capsid protein suggested that these *S*39006 phages represent a new genus of phages. As an increasing number of phage genomes are added to Genbank in recent years, the identification of this new genus of phages will contribute to better understanding of the diversity of T4-family phages (26, 143-145). Characterization of certain known/unknown ORFs of these phages described in later chapters will also add to our understanding of basic phage biology.

Three out of the four *S*39006 phages in this study were not only aborted by ToxIN$_{Pa}$ but also produced spontaneous escape mutants that became insensitive to ToxIN$_{Pa}$. Identification of the escape loci showed that ΦCHI14, ΦCBH8 and ΦX20 could escape ToxIN$_{Pa}$ via three, two and one different route(s), respectively. Of the phage mutants analyzed, the “large deletion” escape route is shared by all three phages while the *asiA* escape route is unique to ΦCHI14 and ΦCBH8. Mutation of ORF84 was only seen in ΦCHI14 and in only one of its mutants. The repeated appearance of similar escape loci of the three *S*39006 phages, and the absence of a “new” escape locus suggests that common themes may exist in terms of how these phages interact with the ToxIN$_{Pa}$ system. This highlights the “large deletion” region and *asiA* gene as important candidates for direct activation of ToxIN$_{Pa}$-mediated Abi that are in need of further investigation.

In contrast, ΦCBH189 was unable to escape ToxIN$_{Pa}$. It is more likely that ΦCBH189 lacks certain proteins required for the genesis of escape mutants, since both *asiA* and the "large deletion" region are almost identical to those found ΦCBH8. Analysis of differences between ORFs of ΦCBH189 and other phages provided little clue since many of the ORFs involved have unknown function. This further complicates the nature of phages that could activate ToxIN$_{Pa}$-mediated Abi, since ΦCBH189 is very closely related to the other three phages but fails to exhibit similar behavior regarding its interaction with ToxIN$_{Pa}$. Future work could involve chemically mutating ΦCBH189, such as using hydroxylamine, to artificially generate escape mutants of ΦCBH189 and analyze their escape loci.

Finally, even though *S*39006 phages in this study were aborted by TenpIN$_{Pl}$, none of them could produce any spontaneous mutants that escaped the system. This might be because the TenpN$_{Pl}$ toxin was overexpressed on a high copy number plasmid, or that the phage gene(s) involved in TenpIN$_{Pl}$-sensitivity is essential, therefore mutants were unable to survive.

In summary, three closely related phages, ΦCHI14, ΦCBH8 and ΦX20, were identified from current and previous studies that contain interesting loci presumed to be involved in activating ToxIN$_{Pa}$-mediated Abi. The following two chapters will focus on two of those loci, the *asiA* gene and the "large deletion" region, to investigate whether they are directly involved in activating Abi and gather information on their possible biological function(s)

Chapter 4 - T4-family phages escape ToxIN$_{Pa}$ via deletion of a large genomic region

1. Introduction

As obligate parasites for their bacterial hosts, phages have a high level of genetic plasticity to adapt to changing host physiology. As phage genome replicates rapidly during the infection process, they can produce missense mutations, nonsense mutations, insertions, inversions, deletions or duplications. Under certain conditions, these spontaneous mutations are selected for and allow phages to quickly adapt to changes, especially to avoid host defense mechanisms. In the case of Type III TA systems, mutants escaping ToxIN$_{Pa}$ and AbiQ have been found where insertion, deletion or missense mutations have enabled them to escape these Abi systems (98, 99). Among ToxIN$_{Pa}$-escape mutants of *S*39006 phages ΦCBH8, ΦCHI14 and ΦX20, deletion of a large genomic region (> 6.5 kb) was observed. Deletions of this extent has not been observed previously for any mutants of Type III TA-sensitive phages. BLAST searches showed that no known function can be assigned to any of the *orfs* deleted in these mutants (Appendix I), making it mysterious as to how these "large deletion" regions are related to ToxIN$_{Pa}$-sensitivity.

This chapter aims at carrying out biological characterization of these "large deletion" regions to answer the following questions: 1) What facilitates the deletion event of regions so large? 2) Are the deleted regions as a whole important for Abi-activation or just single genetic element(s) within the locus? 3) What is the biological relevance of the deleted regions and how are they related to ToxIN$_{Pa}$-sensitivity?

2. "Large deletion" (LD) is the most prevalent mutational route through which *S*39006 phages escape ToxIN$_{Pa}$

Of the 27 sequenced ToxIN$_{Pa}$-escape mutants of ΦCBH8, ΦCHI14 and ΦX20 combined, 21 of them have lost a large region of their genomic DNA (shown in Table 3.3 in the previous chapter).

In particular, all 11 of ΦX20 mutants isolated arose through this "large deletion" (LD) mutation. These "LD mutants" arose for all three phages, therefore making LD the prevalent escape route and the only route through which all these T4-family phages can escape ToxIN$_{Pa}$ (*asiA* mutants were not found for ΦX20). To understand how the deletion of the LD region facilitates escape from ToxIN$_{Pa}$, the locus was characterized in detail.

2.1 The LD region varies between phage mutants but share a common core

Comparison of the sizes of the LD regions in different phage mutants showed that the LD regions range from 6,521 bp (ΦCBH8o) to 10,094 bp (ΦCHI14c). The largest deleted region covers 19 *orfs* and 13 tRNA genes, whereas the smallest covers 13 *orfs* and 11 tRNA genes. The variation in the size of the LD region prompted investigation into whether the LD region could still be smaller so that we could home in on fewer genetic elements to investigate. To look for smaller LD regions, 11 more ToxIN$_{Pa}$-escaping spontaneous mutants of ΦCBH8 were isolated and their genomes extracted. Primers oBH1 and oBH2 designed to flank the LD region in ΦCBH8f (the largest LD region among ΦCBH8 mutants) were used to amplify the cognate regions in these 11 new mutants. The sizes of the amplicons were compared to the size of the amplicon from ΦCBH8o. If any of these 11 mutants had deleted a smaller region than ΦCBH8o, then its amplicon size should be bigger than that of ΦCBH8o. However, visualization of the amplicons on an agarose gel showed that no other mutants have a smaller LD size (data not shown), making ΦCBH8o the mutant with the smallest LD so far.

The exact positions of the deleted region in each mutant were then mapped onto the wild type phage genome to compare similarities and differences. As shown in the schematic diagram in Figure 4.1a-b, no mutation has the exact same 5' or 3' border of deletion as another. The 5' borders of LD mutants seem to be dispersed in a region consisting mostly of tRNA genes, while the 3' borders of LD regions lie in a region full of *orfs* encoding hypothetical proteins. Despite the variability in the borders of the deleted regions, all of the LD regions share a common core that consists of 11 *orfs* and 3 tRNA genes. The location of the LD regions relative to the whole genome was also analyzed by mapping the furthest 5' and 3' borders of all ΦCBH8 and ΦCHI14 mutants combined (64343 bp and 74523 bp). As shown in Figure 3.3, the LD region lies on the forward strand of the phage genome, in a region consisting mostly hypothetical proteins without known function.

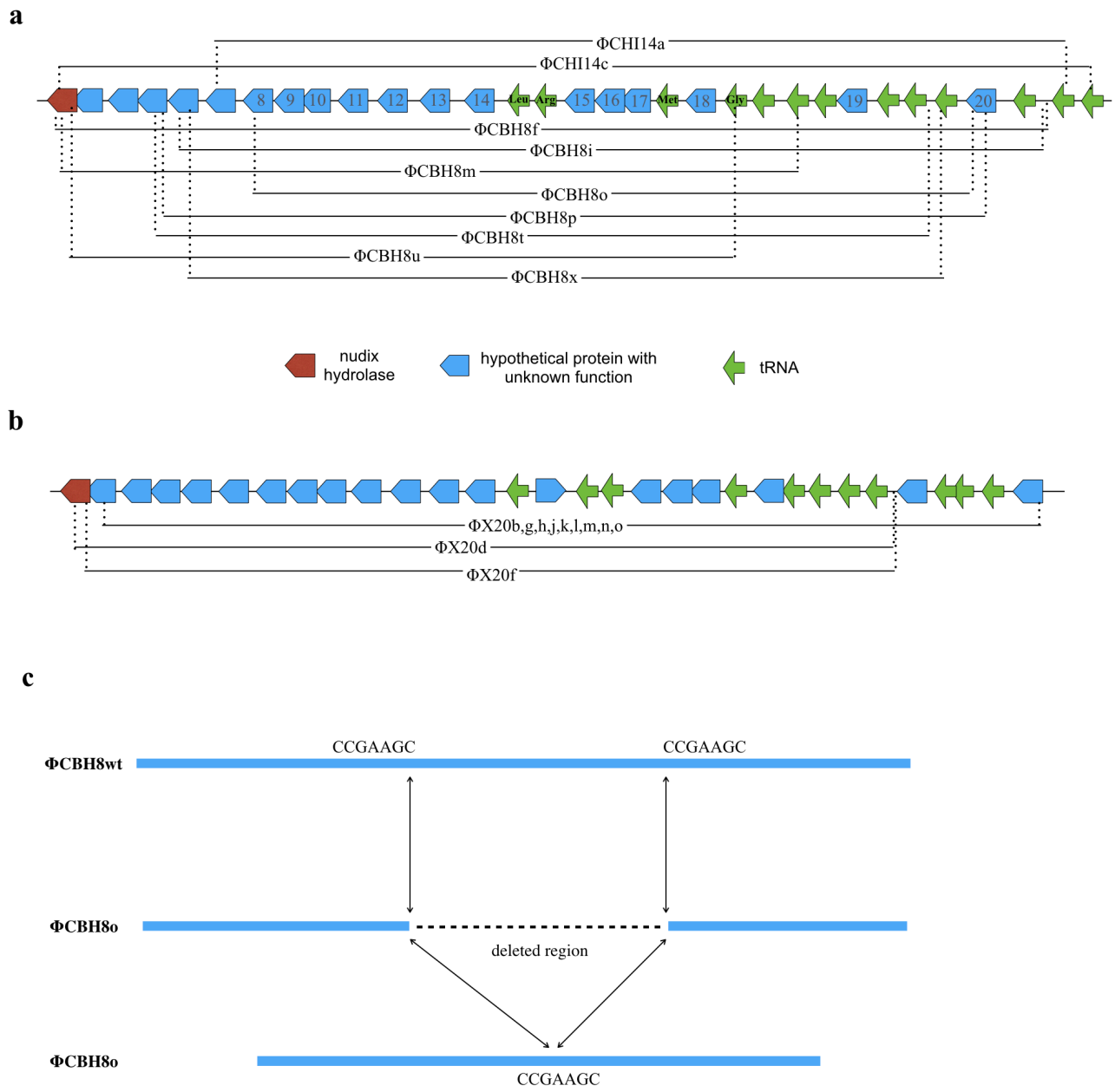


Figure 4.1: Mapping of the deleted regions in different LD. (a) Mapping of the deleted regions in ΦCBH8 and ΦCHI14 LD mutants onto the wild type genome. Due to the 99% similarity between ΦCBH8 and ΦCHI14 wild type genomes, the two are represented as one in the schematic diagram. The numbers mark *orfs* mentioned later in Section 4.4, 4.5 and 4.6 (b) Mapping of the deleted regions in ΦX20 LD mutants onto the wild type genome. The direction of arrows represents the direction of transcription of genes. (c) Schematic diagram showing the existence of direct repeats in the wild type ΦCBH8 genome immediately upstream and downstream of the region deleted in ΦCBH8o, and the cognate region in ΦCBH8o after deletion occurred where only one copy of the repeat sequence is left.

Table 4.1: Information of direct repeats harboring the LD regions in ΦCBH8

Mutant name	LD size (bp)	Direct repeat sequence	Frequency of repeat in wild type ΦCBH8 genome	Frequency of repeat within deleted region
ΦCBH8f	10040	GAACTGC	22	3
ΦCBH8i	7802	TTGAGTAG	7	0
ΦCBH8m	8575	GTCCCTG	11	0
ΦCBH8o	6521	CCGAAGC	15	1
ΦCBH8p	8368	GTTCAC	94	4
ΦCBH8t	7328	AGCCATCC	5	0
ΦCBH8u	8158	GGAAGCC	22	1
ΦCBH8x	7731	ATCTG	>100	11

Investigation of the sequences directly upstream or downstream of the deleted regions showed the presence of direct repeats immediately adjacent to the 5' and 3' borders of the cognate LD regions. These direct repeats, varying from 6 bp to 8 bp in size, are different in each LD mutant (Table 4.1). As exemplified by the case of ΦCBH8o in Figure 4.1c, even though the wild type genome contains 2 copies of the direct repeat sequences adjacent to the region to be deleted, only one copy of the repeat sequence is left after the deletion occurred. This is characteristic of recombination or strand slippage as the mechanism for such deletion (146, 147). Further analysis was also carried out to count the number of times the direct repeats appeared in the ΦCBH8 genome, especially within the deleted regions of mutants. Interestingly, for ΦCBH8o, the mutant with the smallest LD region, the direct repeat also appeared once within the deleted region (Table 4.1). This means that, theoretically, the LD region in ΦCBH8o could have been even smaller than 6.5 kb. However previous screening of more LD mutants showed that 6.5 kb is the smallest deleted region found so far. This could be due to the limited sample size or could imply that a smaller deleted region would not remove *orf*(s) or tRNA(s) that are involved in Abi-activation, therefore such a mutant could not arise as a viable escape mutant during the positive selection.

2.2 The LD region is highly specific to T4-family *S*39006 phages

Alignment of the LD region with the genomes of other related phages was carried out to evaluate whether the region is conserved among T4-family phages. Phages T4, CC31, PEi20, Esc1, ΦCBH189 and ΦX20 were chosen to compare with the LD region deleted in ΦCBH8o in their ascending order of similarity with wild type ΦCBH8 (ΦCBH8wt). As shown in Figure 4.2, alignment of the ΦCBH8o LD region with T4 and CC31 genomes resulted in very low level of identity, and no corresponding LD region exists in these 2 phages. In contrast, ΦCBH8o LD alignment with PEi20 and Esc1 genomes showed moderate levels of identity. The homologous regions in PEi20 and Esc1 are concentrated in one area of the genome, indicating the existence of a cognate LD region (albeit moderate levels of identity) in their genomes. Finally, ΦCBH8o LD alignment with ΦCBH189 and ΦX20 genomes resulted in high level of identity. This is unsurprising since ΦCBH189 and ΦX20 have very high overall genome similarity with ΦCBH8. The corresponding LD regions in ΦCBH189 and ΦX20 are almost identical to ΦCBH8 except for a few gaps. The level of similarity far exceeds the similarity between ΦCBH8o LD and cognate LD regions in PEi20 and Esc1, even though the latter two also have moderately high level of overall genome similarity with ΦCBH8 (Table 3.2, Chapter 3). This shows that as a whole, the LD region presumed to be involved in Abi-activation is highly specific to the *S*39006 phages characterized so far and absent (to different extents) in other related T4-family phages. This result is not surprising since previous results have shown that phage T4, which lacks this LD region, is not able to activate ToxIN$_{Pa}$-mediated Abi in *E.coli* K-12.

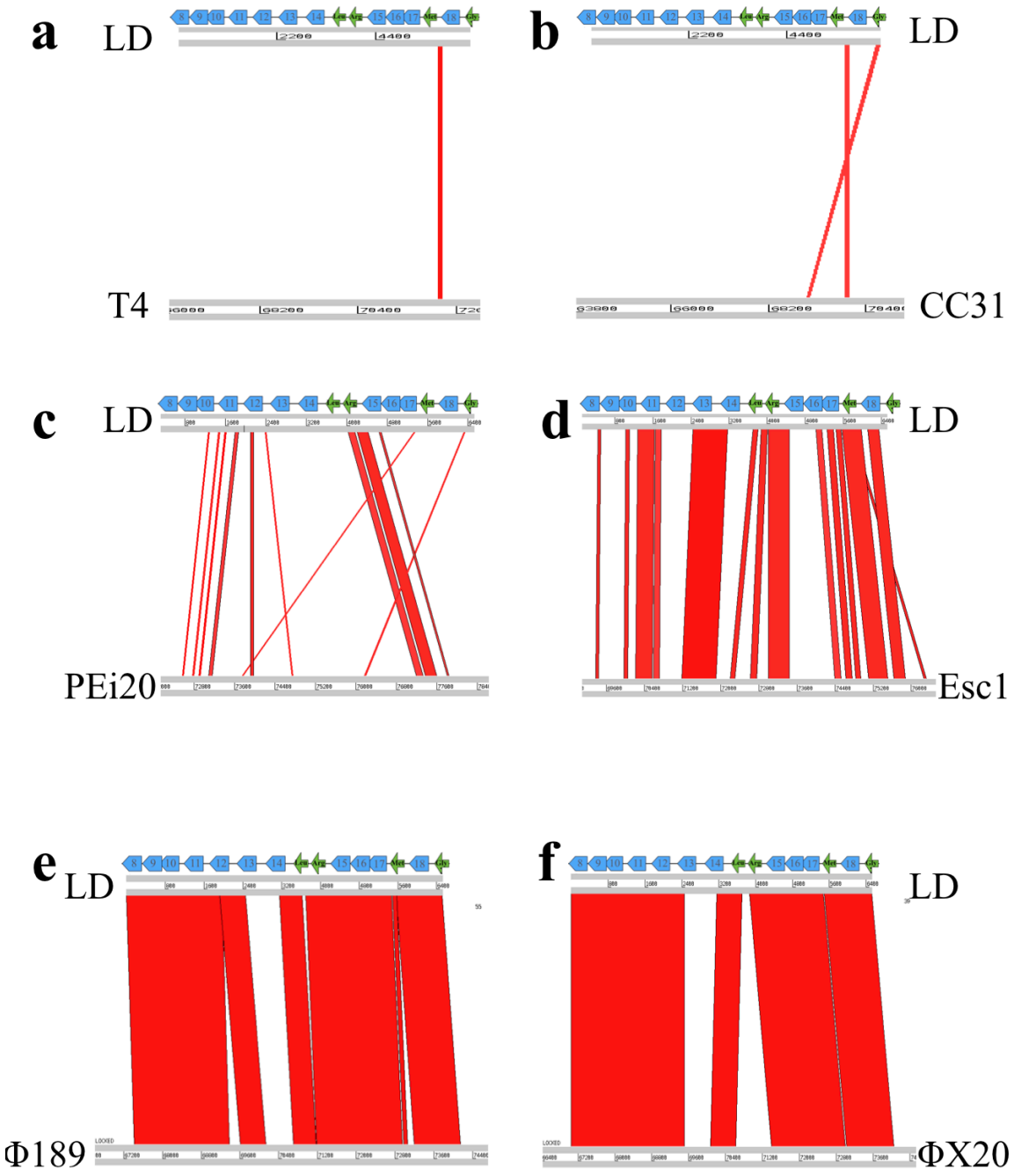


Figure 4.2: ACT alignment of the deleted region in ΦCBH8o (LD) with genomes of highly similar or well-studied phages of the T4 family. (a) phage T4; (b) phage CC31; (c) phage PEi20; (d) phage Esc1; (e) ΦCBH189; (f) ΦX20. While the top panel in each alignment represents the entirety of ΦCBH8o LD, the bottom panel only represents an incomplete portion of the corresponding phage genome that is most similar with ΦCBH8o LD. The red color represents corresponding regions in the alignment that are at least 20 bp in length and exceed 70% sequence identity with each other. Schematic representation of *orfs* in the LD region is imposed on top of the alignments. A clear, detailed version of the annotation can be found in Figure 4.1a.

2.3 The LD region has negligible influence on phage fitness

The apparent specificity of the LD region to *S*39006 phages, and the fact that ToxIN$_{Pa}$-insensitive mutants could afford to lose such a large portion of their genomes to escape yet remain viable, prompted investigation into the importance of the LD region for the fitness of phages such as ΦCBH8. TEM imaging of ΦCBH8o showed that the LD mutant still retained the same morphology as its wild type parent (Figure 4.3a). Measurement of ΦCBH8o showed that ΦCBH8o has a head width of 95 nm (±7 nm), head length of 105 nm (±10 nm), extended tail length of 113 nm (±7 nm), and extended tail width of 19 nm. Accounting for errors, this measurement is in the same range as the measurements for ΦCBH8wt, indicating that the loss of the LD region did not cause observable changes to phage morphology, and implying that the LD region does not contain ORFs involved in phage structure or assembly. The importance of the LD region to plaque morphology was also investigated. As shown in Figure 4.3b, ΦCBH8o plaques formed on a top lawn of *S*39006 are round, clear and with smooth edges, identical to ΦCBH8wt plaques. This suggests that the lytic efficiency of ΦCBH8o is not significantly reduced. The sizes of the plaques are also similar to ΦCBH8wt, corroborating TEM measurements where the size of the ΦCBH8o particle is similar to ΦCBH8wt, since differential plaque size correlates with diffusion speed through agar, stemming from differential phage head size (148).

The fitness of ΦCBH8o was furthered quantified by measuring adsorption efficiency in overnight cultures of *S*39006 at an M.O.I. of 0.01. As shown in Figure 4.4a, in the adsorption assay, the ratio of free (i.e. unadsorbed) phage was measured. The lower the ratio of free phage in the culture, the higher the ratio of adsorbed phages. ΦCBH8wt and ΦCBH8o only had minor differences in the ratio of free phage. Both adsorbed to *S*39006 efficiently, achieving more than 80% phage adsorption by 20 min. This shows that with stationary phase hosts in abundance, ΦCBH8o does not exhibit significantly reduced adsorption efficiency despite lacking the LD region. The burst size of ΦCBH8wt and ΦCBH8o was measured next during one-step growth on exponential-phase *S*39006 (Figure 4.4b). Both ΦCBH8wt and ΦCBH8o have a latent period of 25 min and a rise period of about 20 min. The average burst size was approximately 30 phage particles (pfu; plaque forming units) per initial infection center for ΦCBH8wt and 29 for ΦCBH8o; similar when accounting for errors. Therefore, the above assays indicate that ΦCBH8o does not exhibit obvious reduction in fitness compared to ΦCBH8wt. This implies that the LD region has no fitness impact on phage survival and propagation under these conditions.

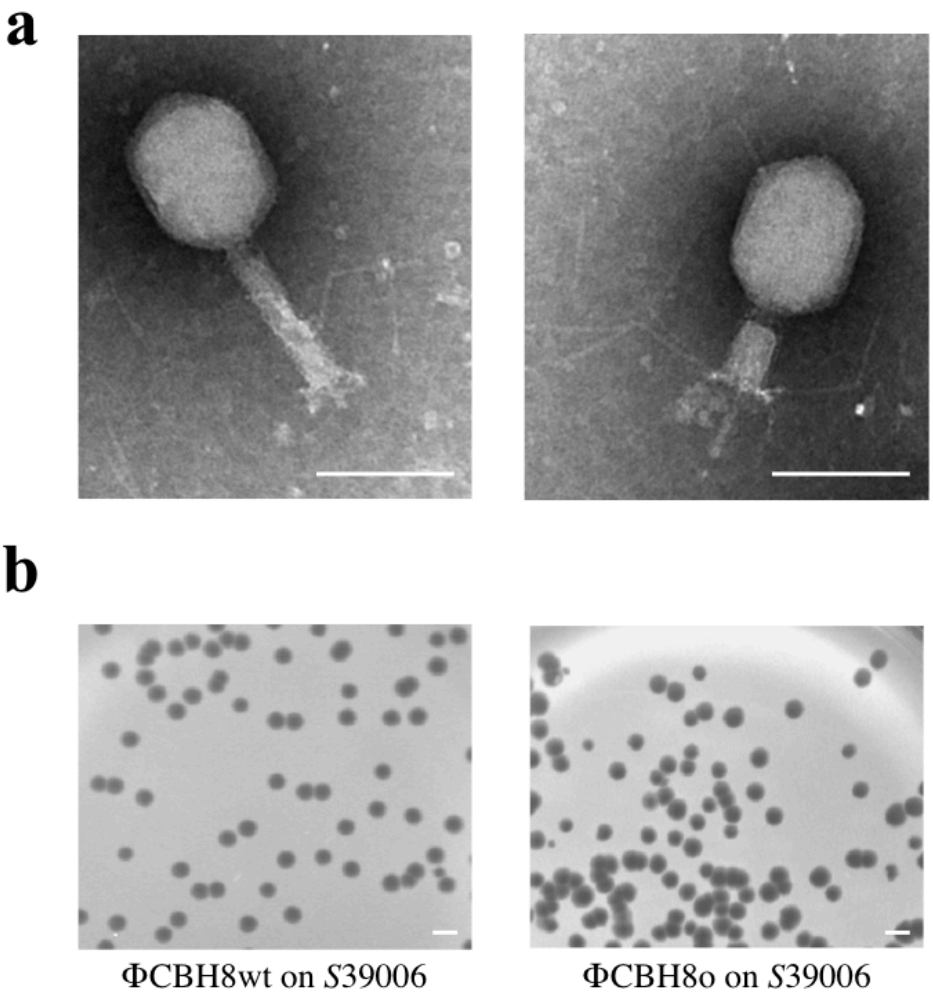


Figure 4.3: The effect of the large deletion (LD) on phage morphology and plaque morphology. (a) TEM images of ΦCBH8o with extended tail (left) and contracted tail (right). Scale bars represent 100 nm. (b) Plaque morphology comparison of ΦCBH8wt (left) and ΦCBH8o (right) grown on top lawn of *S*39006. Scale bars represent 2 mm.

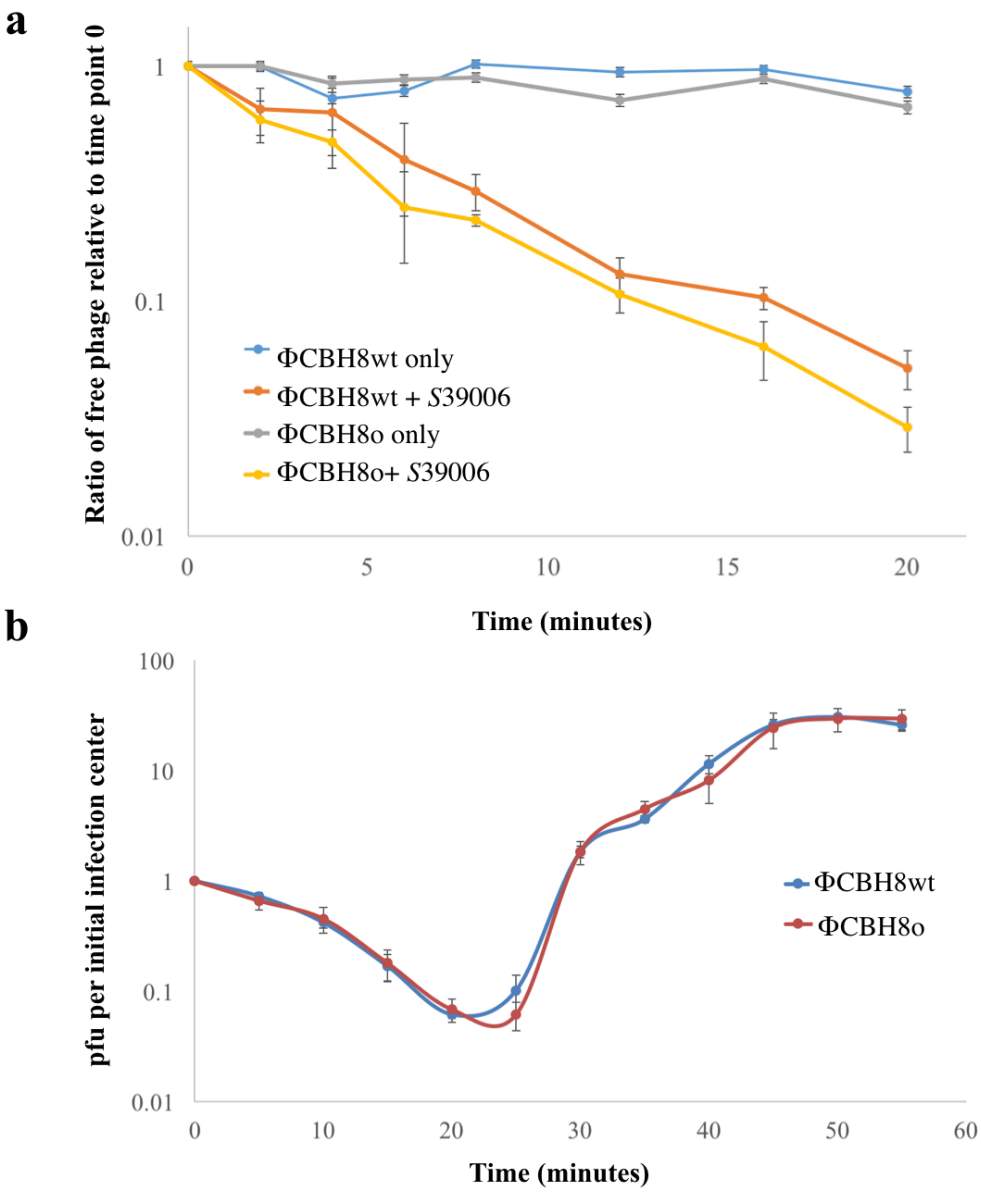


Figure 4.4: Adsorption and burst size comparison of ΦCBH8wt and ΦCBH8o. (a) The adsorption of ΦCBH8wt and ΦCBH8o to *S*39006. Detailed methods of the adsorption assay can be found in Section 3.9, Chapter 2. The percentage of free phage is calculated by dividing the titer of the phage taken from the sample at any given time point by the titer of the phage in the sample at time point 0. The lower the ratio of free phage in the culture, the higher the ratio of adsorbed phages. Error bars represent standard deviation from triplicates. (b) Burst size comparison of ΦCBH8wt and ΦCBH8o infecting *S*39006. Detailed methods of the burst size assay can be found in Section 3.10, Chapter 2. Burst size is calculated by dividing the pfu of phage at any given point by the initial phage titer at time point 0. Error bars represent standard deviation from triplicates.

3. No ORF in the LD region activates ToxIN$_{Pa}$-mediated Abi on its own

The major components of the LD regions are *orfs* encoding hypothetical proteins with no known function. Even in ΦCBH8o, where the deleted region is the smallest, 13 *orfs* were completely eliminated. Similarly, in the overlapping core of all LD regions found in different mutants, there were still 11 *orfs* affected. All the genetic loci identified to be Abi-activating in previous research were shown to be protein-encoding *orfs*, so it is presumed that the *orf*(s) in the LD regions are involved in direct or indirect activation of ToxIN$_{Pa}$-mediated Abi.

To pinpoint which *orf*(s) activates Abi, each one of the 13 *orfs*(denoted *orf*8 to *orf*20) deleted in ΦCBH8o were cloned into pBAD33 and introduced into *S*39006 containing pTA46 (ToxIN$_{Pa}$) for attempts at complementation. The sequence of each cloned *orf* was checked to confirm that none of them have mutated in the cloning process. ΦCBH8o was then serially diluted and spotted on top lawn of *S*39006 expressing both ToxIN$_{Pa}$ and an individual *orf* from the LD region. If any *orf*-encoded product could activate Abi on its own, then ΦCBH8o would be aborted again. As shown in Figure 4.5, with the positive control where ΦCBH8wt was spotted on *S*39006 containing pBAD33 and pTA46, no clear spots were observed, indicating activation of Abi. In contrast, in the negative control (where ΦCBH8o was spotted on *S*39006 containing pBAD33 and pTA46), clear spots were observed. This was expected since ΦCBH8o escaped from ToxIN$_{Pa}$-mediated Abi. However, when *orf*8-*orf*20 (excluding *orf*18) were expressed individually alongside ToxIN$_{Pa}$, none of them caused ΦCBH8o to be aborted again to the extent of the positive control, indicating that none of the *orf* products (excluding ORF18) in the smallest LD region could activate Abi on its own. Expression of *orf*16, *orf*17 and *orf*19 caused minor reduction of phage titer (~10 fold) compared to the negative control, implying that they contribute to Abi to some extent, but must act together with each other or another *orf* to achieve the full effect of Abi.

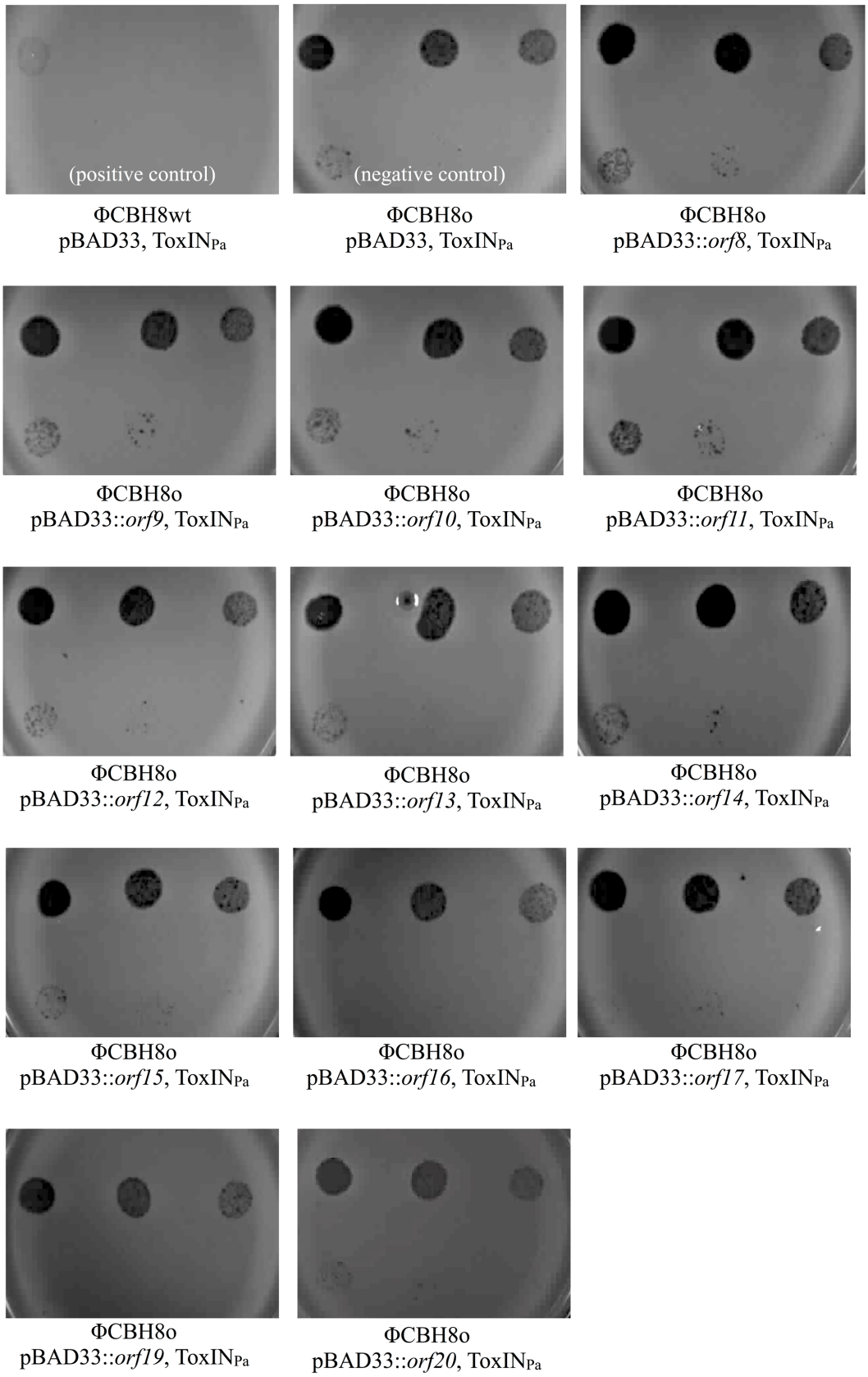


Figure 4.5 (figure legend provided overleaf): Complementation assay of ORFs in ΦCBH8o LD region in the presence of ToxIN$_{Pa}$.

Figure 4.5: Complementation assay of ORFs in ΦCBH8o LD region in the presence of ToxIN$_{Pa}$. All the top lawns contain *S*39006 transformed with ToxIN$_{Pa}$ and an ORF cloned into pBAD33. In positive and negative controls, the empty vector pBAD33 was used. In all the spot tests, 0.05% L-ara was used to induce ORF expression. The undiluted phage stock (10^0 dilution) used had a titer of 10^{10} pfu/mL. Dilution of the ΦCBH8o spots are as follows –
Top row left: 10^{-1}; middle: 10^{-2}; right: 10^{-3}.
Bottom row left: 10^{-4}; middle: 10^{-5}; right: 10^{-6}.

4. ORF18 might be toxic

Attempts to clone *orfs* of the ΦCBH8o LD region into pBAD33 were successful except for *orf* 18 (gene# 146 in Appendix I). This *orf* encodes a hypothetical protein (ORF18) with 114 amino acids. After transformation of DH5α with the ligation mixture of the *orf*18 insert in the pBAD33 vector, either no transformants were obtained on the chloramphenicol selection plate or only 1-2 transformants appeared. Sequencing of the recombinant plasmid in these rare colonies showed that *orf*18 has suffered a missense mutation. The cloning process was repeated several times and even in the presence of 0.4% Glucose for repression, *orf*18 could only be cloned after acquiring missense mutations, suggesting two possibilities: either the cloned portion of the *orf* contains endogenous promoter, or ORF18 is extremely toxic. Search for endogenous promoter of *orf18* based on consensus promoter sequences of T4 genes yielded no result, indicating that either the cloned portion of *orf18* does not have endogenous promoter, or that the promoter is does not conform to the T4 consensus. Whichever scenario applies, the failure in cloning and expression *orf18* explains the failure in attempts to clone and complement the entire LD region of ΦCBH8o (data not shown), within which *orf*18 is contained. Since wild type *orf18* cannot be cloned and expressed, it is not known if *orf*18 can activate Abi on its own. However, given its location on the 5' end of the common core of all LD regions (Figure 4.1a) and its toxicity, it is theoretically possible that its product ORF18 could be involved in Abi-activation.

5. ORF10 causes filamentation in *S*39006

In the absence of direct evidence that any *orf* deleted in ΦCBH8o could activate ToxIN$_{Pa}$-mediated Abi on its own, bioinformatic prediction of the function of the *orf*s was carried out in the hope that some hints could be gained regarding their possible interaction with ToxIN$_{Pa}$. BLASTn and BLASTp searches were carried out to annotate *orf*s in the ΦCBH8o LD region but those *orf*s had either no homologues or only had homologues with no known function. In addition to using sequence information, structural information of these *orf* products was also predicted using i-TASSER (134, 149) to search for potential structural homologues with known function. However only homologues with low Z-scores were returned in the search. Finally, Phagonaute (150) was used to look at the neighboring genes of the homologues of these *orf*s in other phages. Even though some of the ORFs in the ΦCBH8o LD region have high scoring homologues in other phages, those homologues have no known function assigned and were surrounded by hypothetical proteins without known function.

The exhaustive search for functional information of *orf*s8-20 yielded little information, therefore attempts were made to express these *orf*s in *S*39006 to gain some first-hand information on their biological function. Given the fact that these *orf*s were shown to be non-essential for the phage lytic cycle, attempts were made to assess their effect on host fitness. *orf*s8-20 (excluding *orf*18) were cloned into pBAD33 and introduced into wild type *S*39006, individually. A growth curve of *S*39006 expressing each *orf* was plotted where *orf* expression was induced with 0.1% L-ara at the start of exponential phase. The OD_{600} of *S*39006 expressing each *orf* was compared to the control where *S*39006 was transformed with the empty pBAD33 vector. Among the 12 *orfs* assessed, only *orf*10 product ORF10 caused reduction of OD_{600} in cells (Figure 4.6a) whereas the growth curve of cells expressing the rest of the *orf*s was indistinguishable from control (data not shown). Therefore, *orf*10 was singled out and a more refined growth curve was carried out where the colony forming units (cfu) of strains expressing *orf*10 were also counted and the morphology of *S*39006 cells expressing *orf*10 was imaged using phase contrast microscopy. As shown in Figure 4.6b, immediately after induction, cells expressing *orf*10 entered into a bacteriostasis state. The peak cfu of cells expressing *orf*10 after reaching stationary phase was 20-fold lower than cells with the empty vector control. The reason for such reduction could be attributed to stalled cellular division as *orf*10 was shown to cause filamentation of the cells (Figure 4.6c). This could also explain the continued increase in OD_{600} despite the lack of increase in cfu. Compared to control, cells expressing *orf*10 seem to be elongating but not dividing. Some of the elongated cells (Figure 4.6c, pBAD33::*orf10* at 15h) exceeded the length of 10 normal cells, but only 1-2 septa could be observed. This implies that ORF10 does not significantly affect cellular elongation but obstructs septation. This could be due to the innate functionality of ORF10 or could simply be the result of overexpression of an exogenous protein.

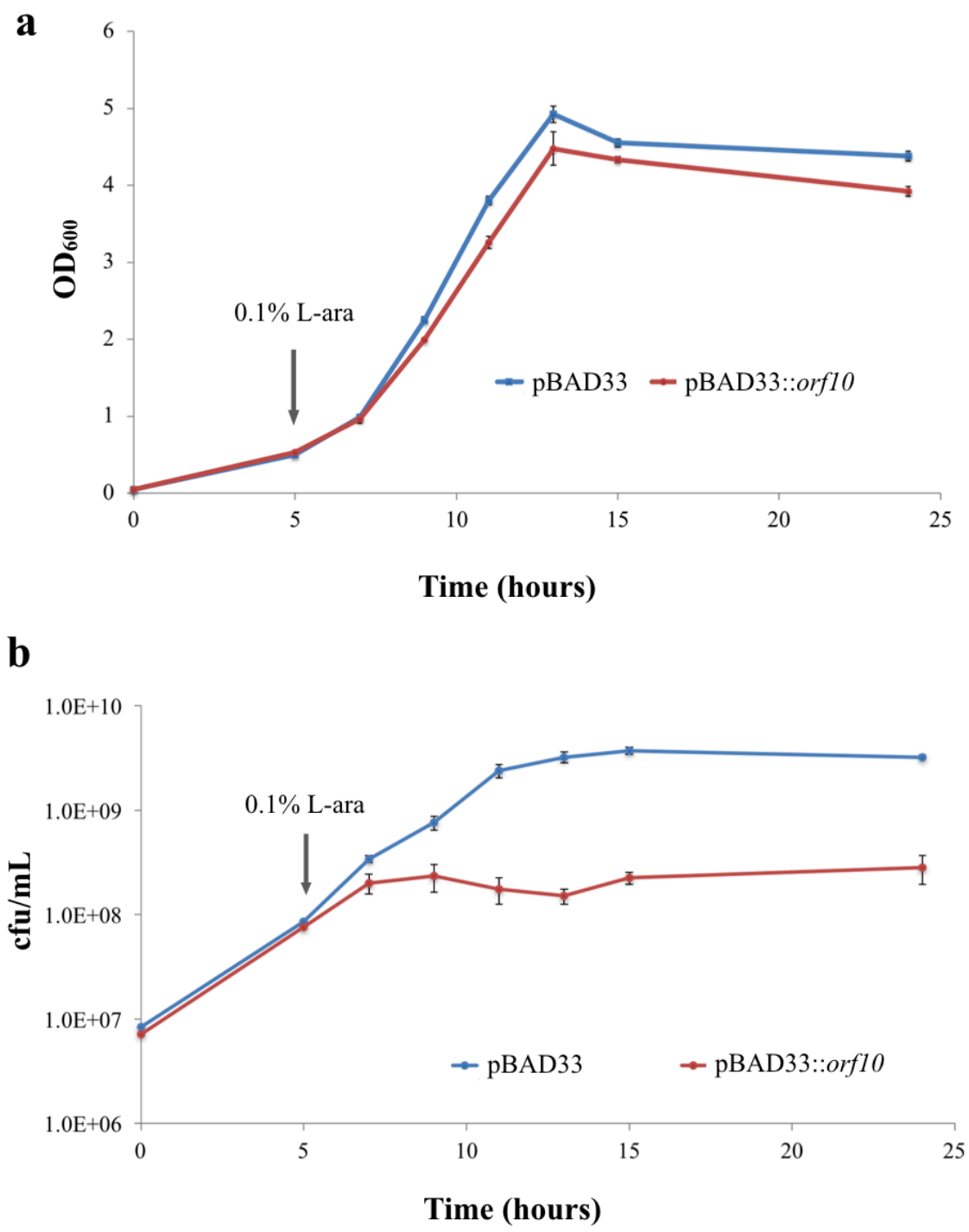


Figure 4.6 (*continued overleaf*): Growth curve of *S*39006 expressing *orf*10.

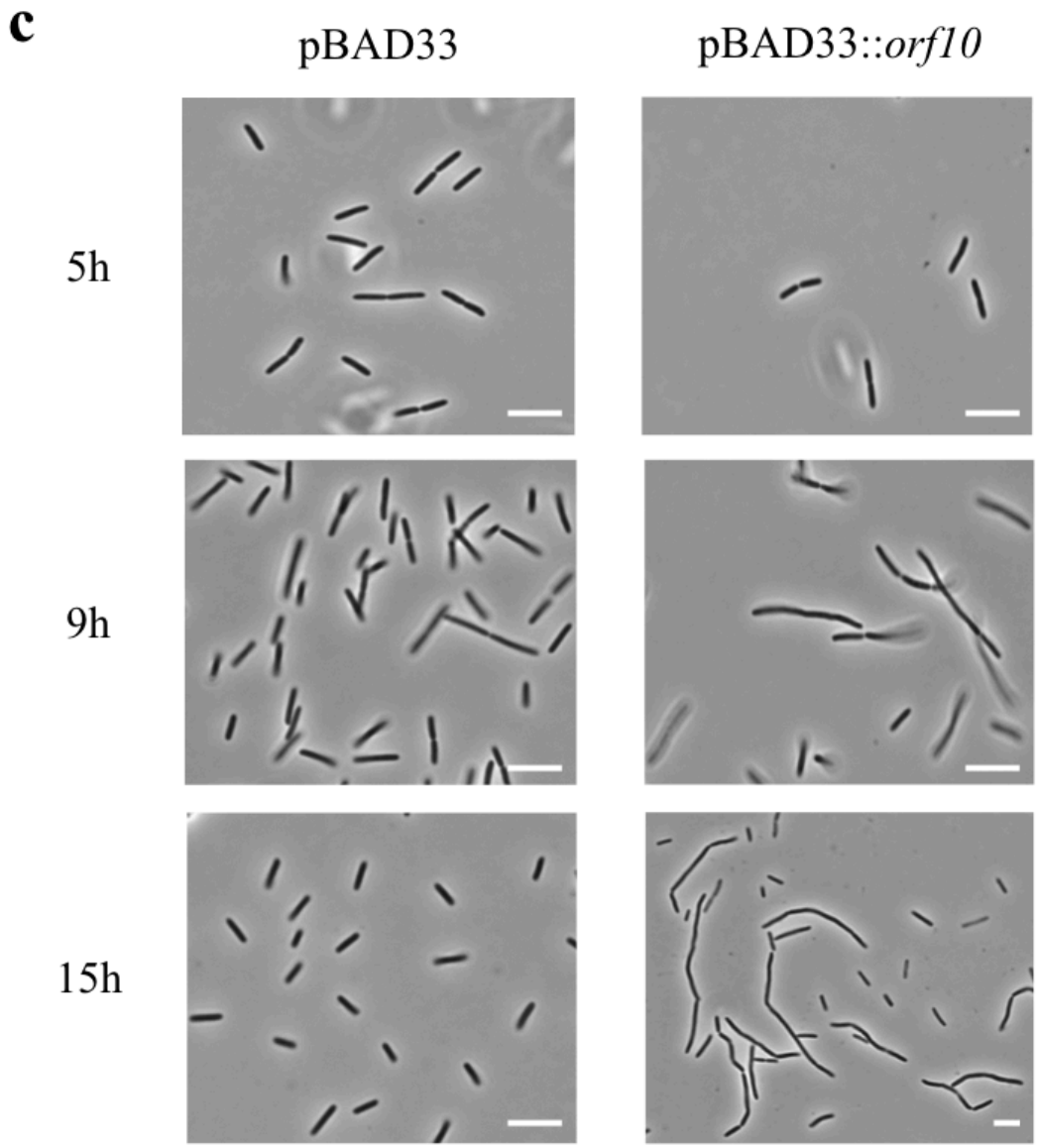


Figure 4.6 (continued from previous page): Growth assay of *S*39006 expressing *orf*10. Detailed method of the growth assay is described in Section 5.1, Chapter 2. The starting culture had an OD_{600} of 0.05 and was grown to early-exponential phase (5h) when 0.1% L-ara was added. (a) Optical density at 600 nm of *S*39006 expressing *orf*10 versus *S*39006 with the empty vector pBAD33. (b) Colony forming units of *S*39006 expressing *orf*10 versus *S*39006 with the empty vector pBAD33. (b) Phase contrast microscopy of *S*39006 expressing *orf*10 versus *S*39006 with the empty vector pBAD33 at hours 5, 9, and 15. Scale bar represents 3μm.

6. No tRNA in the LD region activates Abi on its own

In the LD regions deleted in ToxIN$_{Pa}$-escape mutants of ΦCBH8, there are numerous tRNAs genes. Attempts were made to first investigated whether the tRNAs deleted were responsible for directly activating Abi. ΦCBH8u was chosen as the target mutant because it has the least number of tRNAs contained in its LD region, namely tRNA genes encoding Leu (anticodon TAA), Arg (anticodon TCT), Met (anticodon CAT) and Gly (anticodon TCC) (Figure 4.1). These tRNA genes and surrounding sequences were cloned into pBAD33 individually and introduced into *S*39006 containing ToxIN$_{Pa}$. In a complementation assay, similar to that described in Section 3 of this chapter, induction of tRNA gene transcription with 0.05% L-ara in the presence of ToxIN$_{Pa}$ did not restore Abi of ΦCBH8u (data not shown). This suggests that tRNAs in the LD regions are not directly activating ToxIN$_{Pa}$-mediated Abi, at least individually.

Analysis was also carried out to assess whether the deleted tRNAs could be indirectly activating ToxIN$_{Pa}$-mediated Abi by assisting the translation of *orf* products in the deleted region. Previous research suggested that even though phages could hijack host tRNAs for their own translation, some can still carry tRNA genes in their genomes to complement the different codon usage bias between phage and host (151). Therefore, it is possible that the tRNAs deleted in the escape mutants were important for translation of an Abi-activating phage protein that could not be translated adequately with only host tRNAs. Investigation of the tRNA genes in the *S*39006 genome showed that it contains 1 copy of Leu (TAA), Arg (TCT) and Gly (TCC) genes and 5 copies of Met (CAT). This shows that the host is not completely devoid of the tRNAs deleted in ΦCBH8u. Despite this, the phage may still have a different codon usage profile compared to the host, and a single copy of the above tRNA genes in the host may not have been adequate for efficient translation. Therefore, the codon usage of *orf*s contained in the LD region in ΦCBH8u was analyzed to identify if any ORFs has frequent usage of Leu (TAA), Arg (TCT) and Gly (TCC). As seen in Figure 4.7, the codon usage frequency of all three tRNAs in the phage is higher than in the host, but the difference is minimal. Focusing on *orf*s8-18 within the common core of all deleted regions in different mutants, *orf*10, *orf*16 and *orf*17 have higher codon usage of Leu (TAA) and Gly (TCC) compared to the host or the average of the ΦCBH8 genome, whereas *orf*18 has higher codon usage of Gly (TCC). Despite the relatively higher frequency, the difference in codon

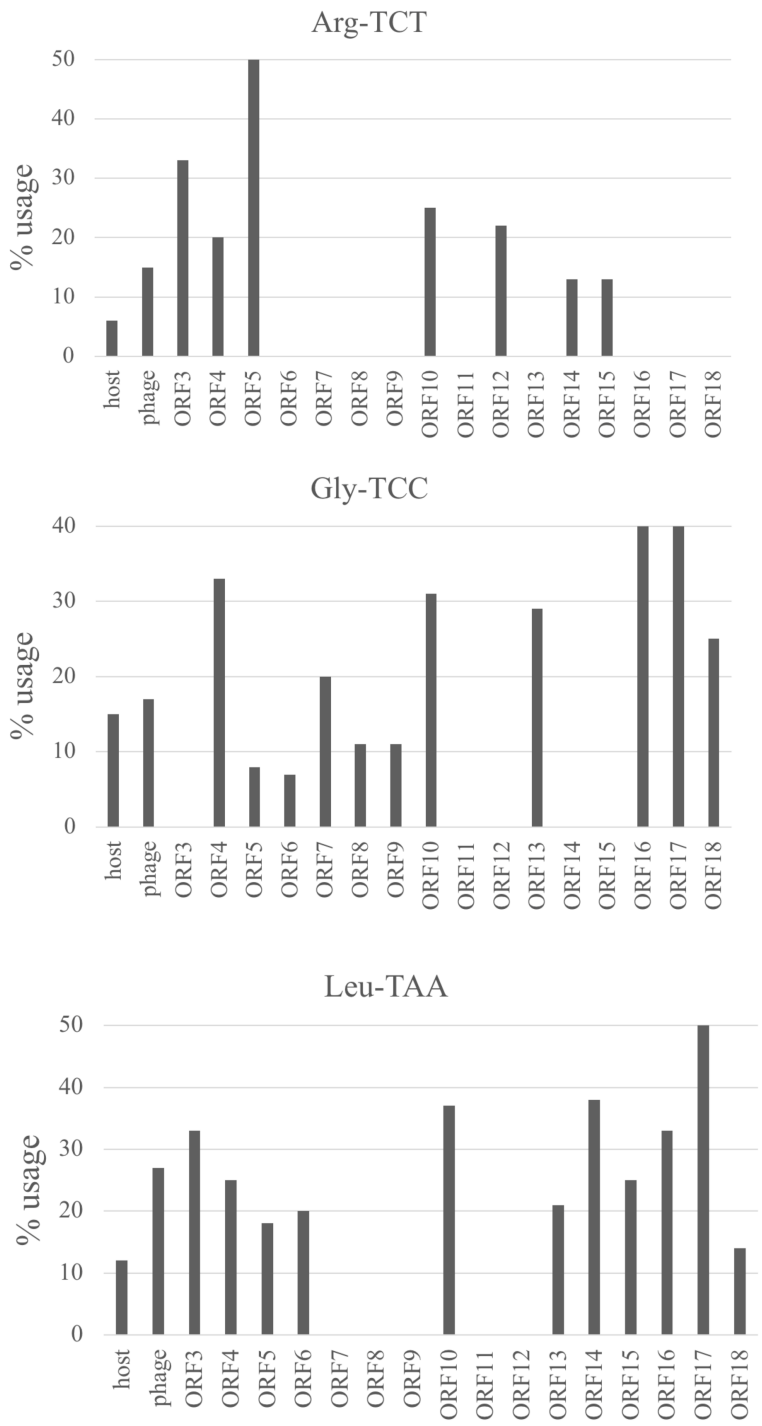


Figure 4.7: Codon usage frequency of Leu (anticodon TAA), Arg (anticodon TCT) and Gly (anticodon TCC). "% frequency" is calculated from the number of times the said codon is used divided by the total number of times that any codon is used for the specific amino acid. "Host" represents the codon usage in *S*39006, "phage" represents the codon usage in ΦCBH8wt, ORF3-ORF18 represent the codon usage of *orf*s deleted in ΦCBH8u.

usages mentioned above are only marginal. Therefore, from the crude bioinformatics analysis, there is no evidence that that deletion of the tRNA genes in ΦCBH8u would affect ORF translation in deleted regions.

7. Discussion

Whole genome sequencing of ToxIN$_{Pa}$-escape mutants of ΦCBH8, ΦCHI14 and ΦX20 revealed that the most frequently found escape locus was the deletion of a substantial region of the phage genome, presumably indicating that elements within these regions activate ToxIN$_{Pa}$-mediated Abi. The deletions may have occurred through recombination or strand slippage, as suggested by the existence of direct repeats bordering the deleted regions. The size of the deletion represented 3.8–5.8% of the total genome size, yet no obvious fitness defects were observed for the mutants, confirming that the deleted regions are dispensable for phage propagation under lab conditions. This is consistent with previous research showing that some phage genomes carry genes that are not necessary for propagation, at least in the growth conditions tested (152, 153). However, this does not mean that the deleted regions are completely redundant, they may have functions under unfavorable conditions to improve phage fitness, such as in the natural environment, or perhaps to trigger host defense. This is intriguing in that a "redundant" part of the phage genome acts against the phage itself to activate an Abi system. One explanation for this would be that these *S*39006 phages are not usually exposed to ToxIN$_{Pa}$ in nature, and the LD regions' involvement in Abi activation is mere coincidence or side effect. Another explanation would be that the entire LD region is acquired by wild type phages through recombinogenic lateral gene transfer (154) in the natural environment (perhaps during mixed viral infections). This is supported by two pieces of evidence: (1) the LD region is unique to these ToxIN$_{Pa}$-sensitive *S*39006 phages and has limited degree of identity with other related phages; and (2) most of the 5' borders of the deleted regions lie within tRNA genes, whose conserved sequences are readily recombinogenic. It could be argued that this has echoes of the descriptions of Pathogenicity Islands in bacteria (155, 156) and therefore could even suggest that the large deletion region is a potentially mobile viral genome unit.

Further investigation into the genetic elements contained in the deleted regions showed the presence of protein-coding *orf*s and several tRNA genes. It is more likely that the *orf*s, instead of tRNAs, are responsible for Abi-activation for 4 reasons: (1) genetic loci involved in Type III-TA activation in other systems were all found to be protein-coding *orf*s; (2) the 4 tRNAs deleted in

ΦCBH8u were present in the host genome; (3) for the 4 tRNAs deleted in ΦCBH8u, no significant codon usage bias difference was observed between the host and phage, or between *orfs* in the LD region and host and (4) complementation of the deleted tRNA genes did not "re-activate" Abi during ΦCBH8u infection. *orf*s in ΦCBH8o, the mutant that contains the least number of *orf*s deleted, were investigated one by one. Complementation assays showed that none of the *orf*s could individually activate ToxIN$_{Pa}$, suggesting that at least 2 *orf*s are simultaneously involved in the process. These *orf*s could be located some distance apart from each other, so that it could be easier to delete the entire region of non-essential genes containing these *orf*s rather than mutating each of them simultaneously. Evidence corroborating this hypothesis comes from the toxicity of the *orf*18 product and obstruction of cell division from the *orf*10 product. At the same time, *orf*18 and *orf*10 are located at two extremes of the common core shared by all LD regions (Figure 4.1). Therefore, *orf*10 and *orf*18 are two candidates that may be involved in activating ToxIN$_{Pa}$-mediated Abi in concert. Given the possible toxic nature of *orf*18, future research could focus on deleting *orf*10 and *orf*18 from ΦCBH8wt and assessing the resulting ToxIN$_{Pa}$-sensitivity. In addition, ORF16, ORF17 and ORF19 were found to cause partial reduction of ΦCBH8o titer in the complementation assay, implying that they could also be involved in activating Abi. Therefore future attempts could also involve investigation of any direct interactions between these ORFs and the ToxIN$_{Pa}$ complex using biochemical techniques.

In conclusion, characterization of the "large deletion" mutants in this chapter have presented a frustrating and complicated scenario where more than two *orf*s are likely to be simultaneously involved in Abi-activation. This is different from previous examples of ToxIN$_{Pa}$ or AbiQ escape mutants where only a single *orf* was mutated (98, 99). If indeed *orf*10 and *orf*18 are involved in activating ToxIN$_{Pa}$-mediated Abi, then a similar feature with *orf*23 in ΦM1 is that both encode a toxic, but unrelated, hypothetical protein. These *orf*s could be triggering some host stress-response pathway that, in turn, could cause Abi activation. Future work could test this hypothesis by looking at ToxIN$_{Pa}$-mediated Abi activation in hosts with defective stress-response systems.

Chapter 5 - T4-family phages escape ToxIN$_{Pa}$ via mutation of the *asiA* gene

1. Introduction

For virulent phages such as T4, transcription of its own genes happens rapidly after DNA injection. In the absence of its own RNA polymerase (RNAP), T4 must hijack the host transcription machinery. The key process in hijacking involves redirecting the σ^{70} subunit of host RNAP away from host promoters and towards phage promoters. While promoters of early T4 genes have naturally high affinity for host σ^{70}, the mechanism for phage middle gene transcription is much more complicated (26). The transcription of some middle genes occurs through a process called σ-appropriation where RNAP is redirected towards phage middle promoters through the concerted effort of a phage-encoded transcription activator MotA and a co-activator AsiA. As reviewed extensively by Hinton and colleagues and illustrated in the simplified diagram in Figure 5.1, AsiA re-models σ^{70} through its multiple contacts with Region 4 of σ^{70} and β-flap of the RNAP β-subunit (25, 157). This remodeling retards RNAP's ability to recognize -35 elements of host promoters and phage early promoters and frees the C-terminal domain (CTD) of σ^{70} so that it can form contact with the CTD of MotA. In the meantime, the N-terminal domain (NTD) of MotA recognizes and binds to the MotA box, a phage middle gene-specific element located at -30. This enables host RNAP to recognize phage middle gene promoters and initiate transcription. In this complicated process, AsiA can be seen as a bridge that connects σ^{70}, the β-flap and MotA (25, 158).

As briefly introduced in Chapter 3, *S*39006 phages ΦCBH8 and ΦCHI14 could escape ToxIN$_{Pa}$-mediated Abi through mutation of the *asiA* gene. In contrast to the situation with hypothetical proteins in the LD regions of other *S*39006 phages or protein M1-23 in ΦM1 (98), this is the first case where an ORF associated with ToxIN$_{Pa}$-mediated Abi has known function. However, given the important yet complicated role that AsiA plays in σ-appropriation, it is not immediately clear how AsiA is connected to the activation of Abi.

Therefore, this chapter aims to characterize the effect of the *asiA* mutations on the phage, the host, and ToxIN$_{Pa}$. Phage fitness in the presence of wild type *asiA* gene and mutant *asiA* gene will be

tested. The effect of producing the wild type AsiA protein and mutant AsiA protein on *S*39006 growth will also be assessed. Finally, the effect of expressing the wild type and mutant *asiA* gene on the activity of the *toxIN$_{Pa}$* locus will be investigated using a reporter system. These assays will provide an answer to whether the *asiA* gene's involvement in ToxIN$_{Pa}$ activation occurs through direct or indirect ways.

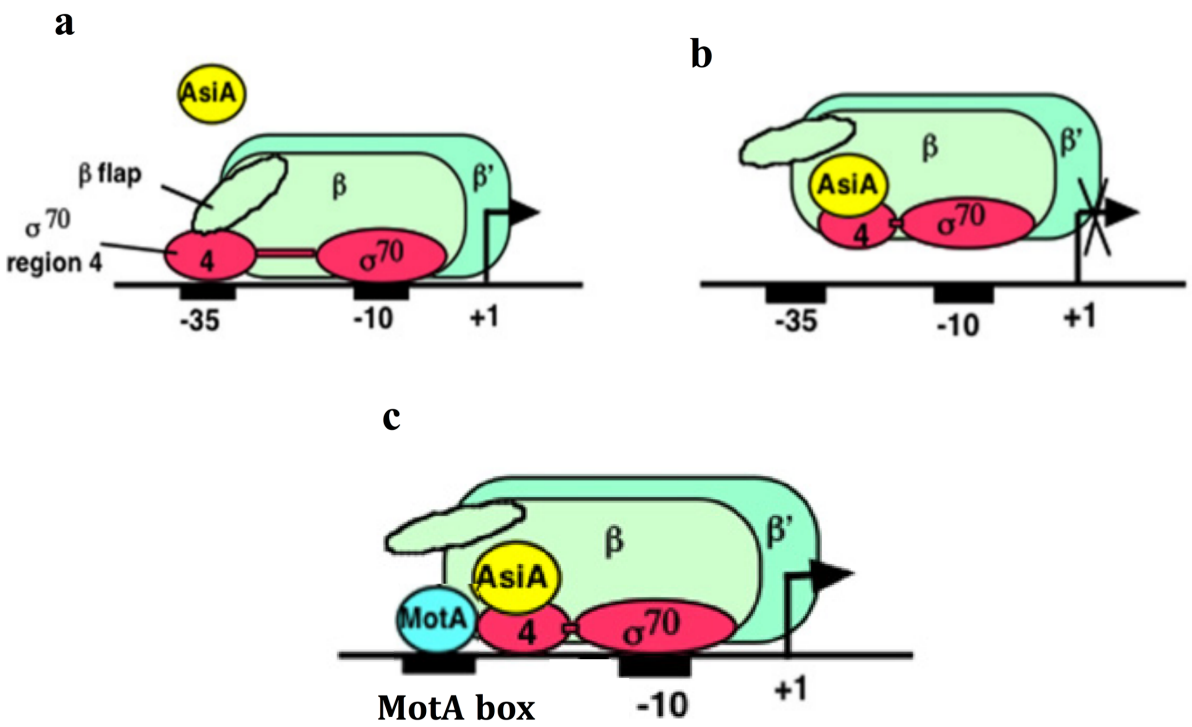


Figure 5.1: The process of σ-appropriation in T4. (a) Without AsiA, RNAP transcribes host genes by recognizing promoter -35 elements through Region 4 of σ^{70}. (b) After the AsiA protein, produced by T4 early gene *asiA*, contacts σ^{70} Region 4 and β-flap, σ^{70} undergoes conformational change and is no longer able to recognize the promoter -35 element, and is thus unable to transcribe phage early genes or host genes. (c) The conformational change caused by AsiA also enables σ^{70} to form contact with MotA, which recognizes and binds to the MotA box at phage middle promoters. This enables RNAP to start transcribing phage middle genes. In all the figures, α-subunits of RNAP are not shown for simplicity. Figures are adapted from the research article by Minakhin and colleagues (159).

2. AsiA is involved in the activation of ToxIN$_{Pa}$-mediated Abi

ToxIN$_{Pa}$-escape mutants of ΦCBH8 and ΦCHI14 (ΦCBH8a, ΦCBH8l, ΦCBH8r, ΦCHI14b and ΦCHI14f), were found to have mutations in the *asiA* gene based on whole genome sequencing data. To validate the mutations, the *asiA* genes in these mutants were PCR amplified and subjected to Sanger sequencing. Furthermore, additional ToxIN$_{Pa}$-escape mutants of ΦCBH8 and ΦCHI14 were isolated in an attempt to find more mutants with *asiA* mutations. This resulted in the isolation of three more mutants (ΦCHI14q, ΦCHI14s and ΦCHI14w) (105). In total, 8 *asiA* mutants of ΦCBH8 and ΦCHI14 were sequenced and the nature of the mutations shown in Figure 5.2. Most of the mutations cause changes at the CTD of AsiA except for ΦCHI14w where the Valine to Glycine mutation is near the NTD. At the CTD, the mutated AsiA is either truncated after the 70th or 71st amino acid or undergoes amino acid substitution followed by extension after the 80th or 81st amino acid.

To further confirm that AsiA plays an important role in enabling ΦCBH8 and ΦCHI14 to activate ToxIN$_{Pa}$-mediated Abi, complementation of AsiA was carried out. The *asiA* gene from wild type ΦCBH8 was PCR amplified and cloned into the L-ara inducible pBAD33 vector. ΦCBH8a (CTD truncated AsiA) was used to infect *S*39006 transformed with both ToxIN$_{Pa}$ and pBAD33::*asiA*. As shown in Figure 5.3, when expression of wild type AsiA was not induced, ΦCBH8a formed clear spots on the bacterial top lawn, indicating no activation of Abi. In contrast, when expression of wild type AsiA was induced with 0.05% L-ara, ΦCBH8a was not able to lyse the bacterial lawn, indicating Abi. The complementation assay confirms that the presence of wild type AsiA is indeed essential for activating ToxIN$_{Pa}$-mediated Abi.

```
CBH8wt   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14wt  1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14b   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14f   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14q   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14s   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14w   1 MSKIEIVREIVTGASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8a    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8l    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8r    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60


CBH8wt  61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA----------------------------- 89
CHI14wt 61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA----------------------------- 89
CHI14b  61 LNDDERSSLV-------------------------------------------------- 70
CHI14f  61 LNDDERSSLVEEFNEGFEDIWLCITRKGDRSLLILLFLKYGRPVDI------------ 106
CHI14q  61 LNDDERSSLVEEFNEGFEDISIVIWLCITRKGDRSLLILLFLKYGRPVDI--------- 110
CHI14s  61 LNDDERSSLVEEFNEGFEDIYIVIWLCITRKGDRSLLILLFLKYGRPVDI--------- 110
CHI14w  61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA----------------------------- 89
CBH8a   61 LNDDERSSLV-------------------------------------------------- 70
CBH8l   61 LNDDERSSLVE------------------------------------------------- 71
CBH8r   61 LNDDERSSLVE------------------------------------------------- 71
```

Figure 5.2: Alignment of the AsiA protein sequence in wild type ΦCBH8 (CBH8wt) and ΦCHI14 (CHI14wt) with mutated AsiA protein in ToxIN$_{Pa}$-escaping mutants (CHI14b, CHI14f, CHI14q, CHI14s, CHI14w, CBH8a, CBH8l, CBH8r). Note that AsiA sequences from wild type ΦCBH8 and ΦCHI14 are identical. Blue highlight indicates conserved residues in mutants compared to wild type while yellow highlight represents mutated residues, including single amino acid substitution (CHI14w) or frameshift mutation causing truncation (CHI14b, CBH8a, CBH8l, CBH8r) or substitution followed by extension (CHI14f, CHI14q, CHI14s).

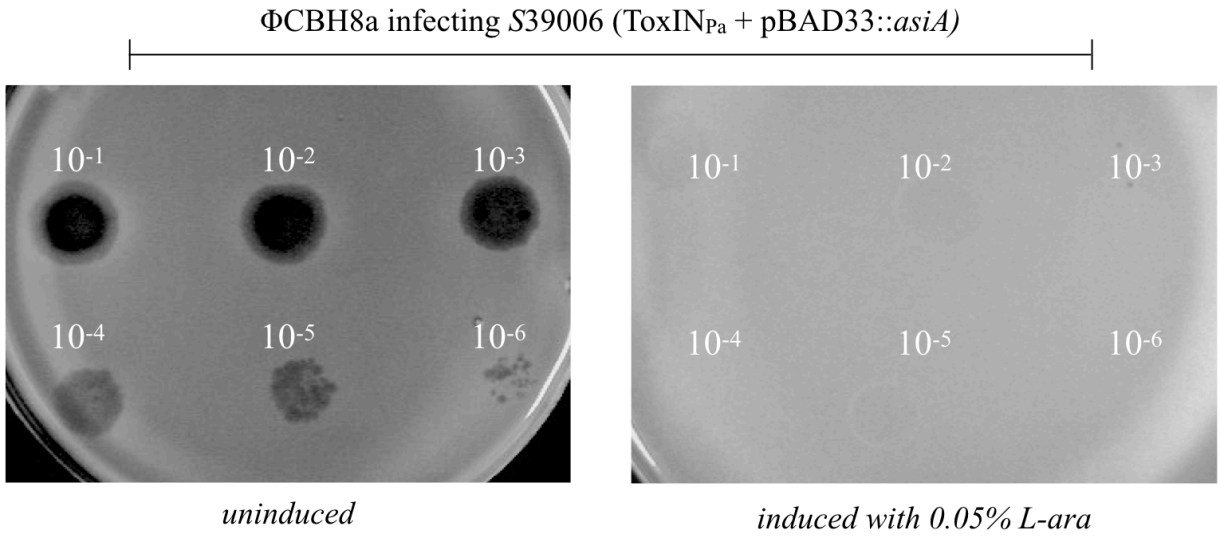


Figure 5.3: Complementation of AsiA. ΦCBH8a was serially diluted and plated on a top lawn containing *S*39006 transformed with ToxIN$_{Pa}$ and pBAD33::*asiA*. When *asiA* expression was not induced (left), ΦCBH8a formed clear spots and plaques on the bacterial lawn, indicating escape from ToxIN$_{Pa}$. When *asiA* expression was induced with 0.05% L-ara (right), ΦCBH8a did not cause lysis of the bacteria even at high phage titer, indicating Abi. In the experiments above, the undiluted phage stock (10^0 dilution) used had a titer of 10^{10} pfu/mL.

3. AsiA in ToxIN$_{Pa}$-escape mutants exhibit loss of functionality

To provide more context for AsiA mutations found in ToxIN$_{Pa}$-escape mutants of ΦCBH8 and ΦCHI14, the AsiA primary sequences from different T4-family phages mentioned in Table 3.2 (Chapter 3) were aligned with each other. As shown in Figure 5.4, ΦCBH8 and ΦCHI14 have identical AsiA consisting of 89 amino acids whereas AsiA sequences from other T4-family phages are 1-7 amino acids longer. Amino acid residues that are fully conserved or have conservative replacements (highlighted in red in Figure 5.4) are evenly distributed throughout the lengths of the peptides. In particular, the NTD and CTD showed greater continuity of sequence conservation, as well as the 33rd -39th residue region (in reference to T4 AsiA). The mutated Valine in ΦCHI14w is positioned in a fully conserved region in the NTD and the truncated or substituted and extended regions in the CTD of other mutants also contain conserved residues. Since sequence conservation is often correlated with functional conservation, it is possible that the mutated protein product of *asiA* in ToxIN$_{Pa}$-escape mutants have compromised functionality.

To investigate whether the mutated AsiA protein from ToxIN$_{Pa}$-escape mutants was still functional, a two-pronged approached was used where the effect of *asiA* mutations on both the host and phage fitness was investigated. On the host side, a growth assay of *S*39006 transformed with the pBAD33 vector carrying different versions of *asiA* was conducted. As shown in Figure 5.5a, *S*39006 carrying the empty pBAD33 vector (negative control) was grown in LB, and culture OD_{600} exhibited the conventional S-shaped curve and reached ~4.5 after entering stationary phase. Cells also became smaller and shorter upon reaching stationary phase (Figure 5.5b), which is characteristic of reductive division and dwarfing for stationary phase cells (160). In contrast, where *S*39006 was transformed with the wild type AsiA (positive control), expression of *asiA* caused a reduced growth rate in culture OD_{600}, and when cells entered stationary phase, culture OD_{600} was less than half of the negative control. Cell morphology did not exhibit any obvious changes after entering stationary phase. This suggests that expression of wild type *asiA* of ΦCBH8 slowed down *S*39006 cell division and eventually caused growth arrest. This is expected since the function of AsiA in T4 is prevention of host gene transcription, leading to diminished growth and growth arrest. Interestingly, in the test cultures where *S*39006 expressed the mutated AsiA protein from ΦCHI14w (V13G) or ΦCBH8a (CTD truncation), cells did not exhibit any growth arrest and behaved almost identical to the negative control. This implies that mutated AsiA can no longer inhibit host

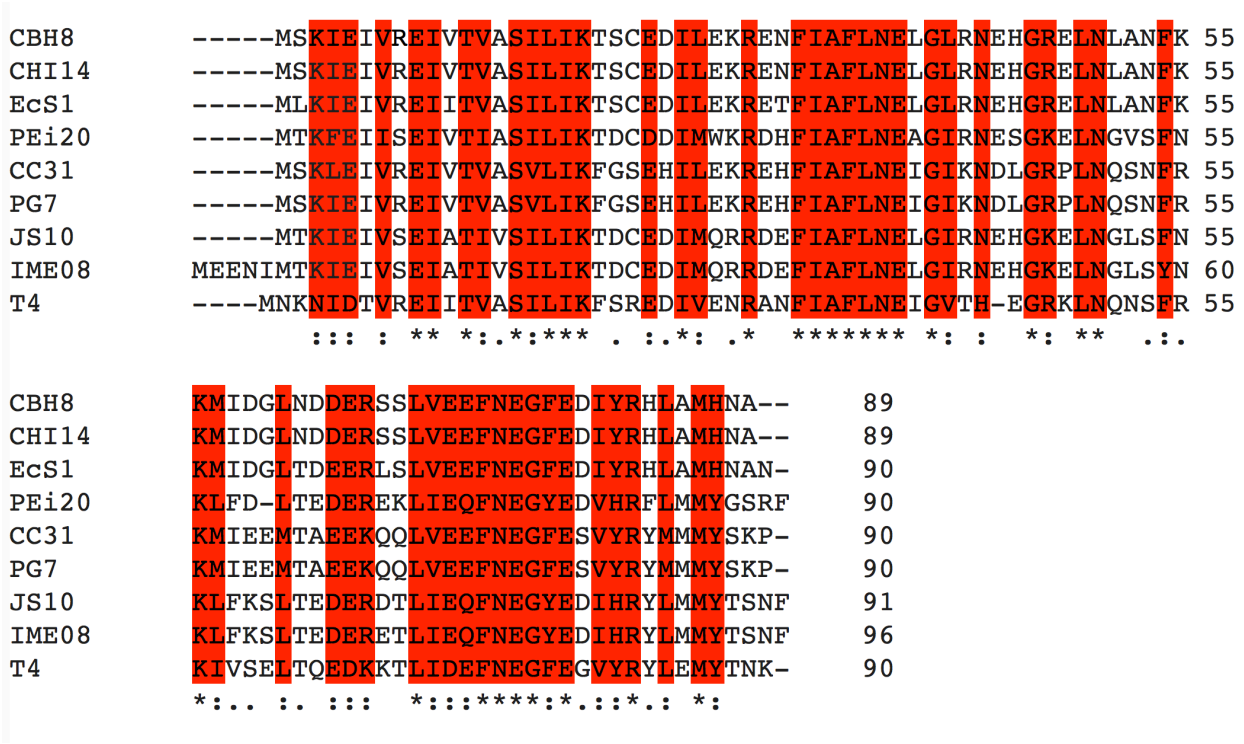


Figure 5.4: Alignment of AsiA protein in T4-family phages. "*" (asterisk): positions that have a single, fully conserved residue. ":" (colon): conservation between groups of strongly similar properties. "." (period): conservation between groups of weakly similar properties. Positions that have either fully conserved or strongly similar amino acid residues are colored in red.

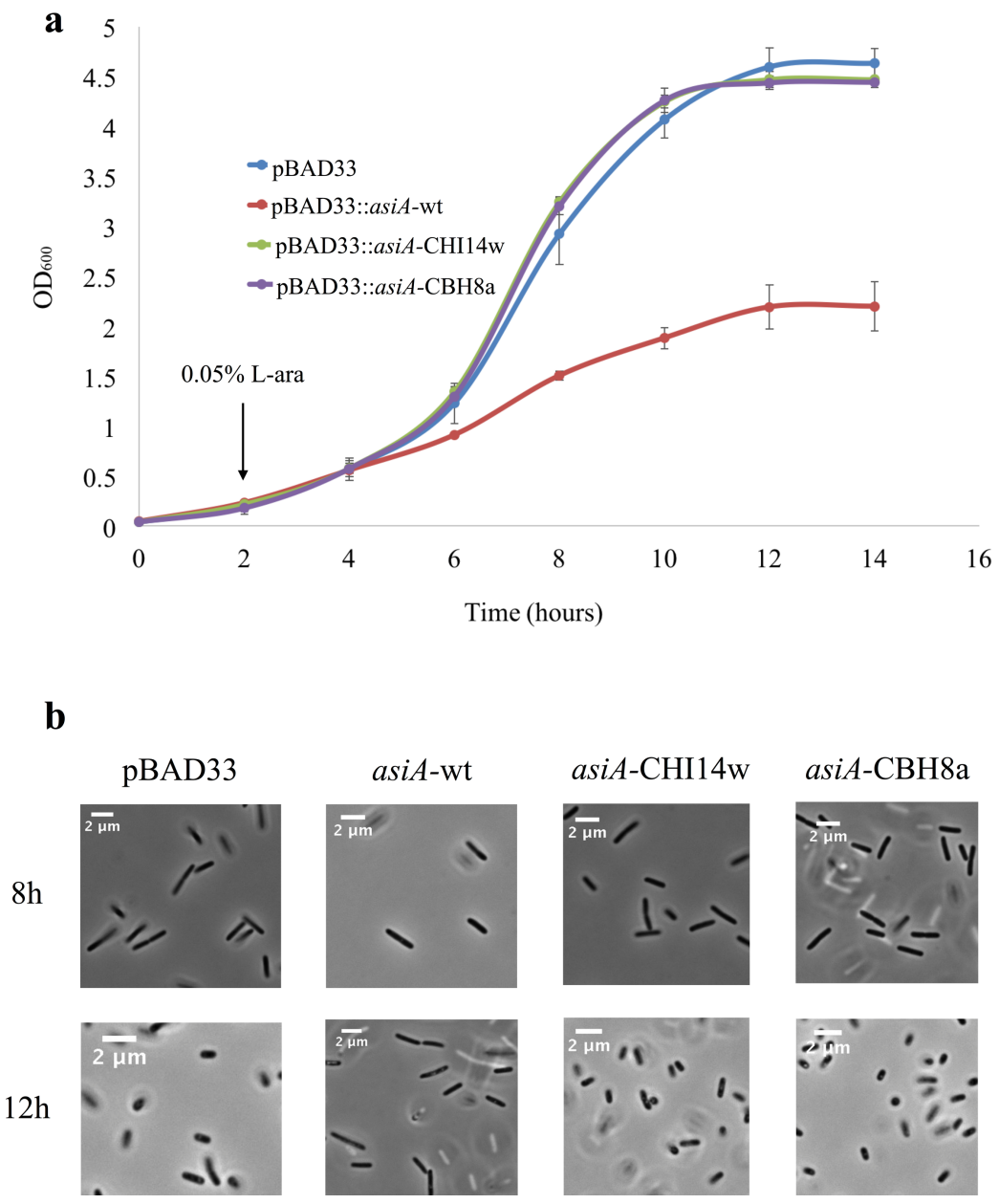


Figure 5.5: *S*39006 expressing wild type or mutant AsiA. In the negative control, *S*39006 was transformed with the empty pBAD33. In the positive control, *S*39006 was transformed with the wild type *asiA* cloned into pBAD33 (pBAD33::*asiA*-wt). In the test samples, *S*39006 was transformed with *asiA* from ΦCHI14w or ΦCBH8a cloned into pBAD33 (pBAD33::*asiA*-CHI14w or pBAD33::*asiA*-CBH8a). Cells were grown from a starting OD_{600} of 0.05. Induction of *asiA* expression was achieved by adding 0.05% L-ara at 2h. (a) OD_{600} of the cultures, taken every two hours until stationary phase. Error bars represent standard deviation from triplicates. (b) Phase-contrast microscopy showing cellular morphologies at 8h (exponential phase) and 12h (early stationary phase). Scale bars represent 2 µm.

gene expression, consistent with the mutations causing loss of function. The different positions of the mutations also suggest that both V13 and the C-terminus residues are important for a functional AsiA.

On the phage side, if mutated AsiA protein lost its function in inhibiting host cell growth, it is likely that phages with mutant AsiA protein would also have fitness defects because successful phage replication is dependent on the function of AsiA in σ-appropriation. The plaque size and morphology of *asiA* mutant ΦCBH8a was first investigated in comparison to ΦCBH8wt. As shown in Figure 5.6a, when grown on a top lawn of wild type *S*39006, ΦCBH8a formed pinpoint plaques. Since plaque size is considered to be positively correlated with burst size for some phages (161), this could imply that ΦCBH8a has a decreased burst size when infecting wild type *S*39006. Indeed, investigation of ΦCBH8a burst size showed that ΦCBH8a has significantly decreased burst size (~10 phage particles per center of infection), in comparison to the ΦCBH8wt burst size of 30 (Figure 5.6b). Furthermore, ΦCBH8a also had a significantly longer latent period. This means that mutation of AsiA in ToxIN$_{Pa}$-escape mutants seriously compromises the efficiency of phage propagation.

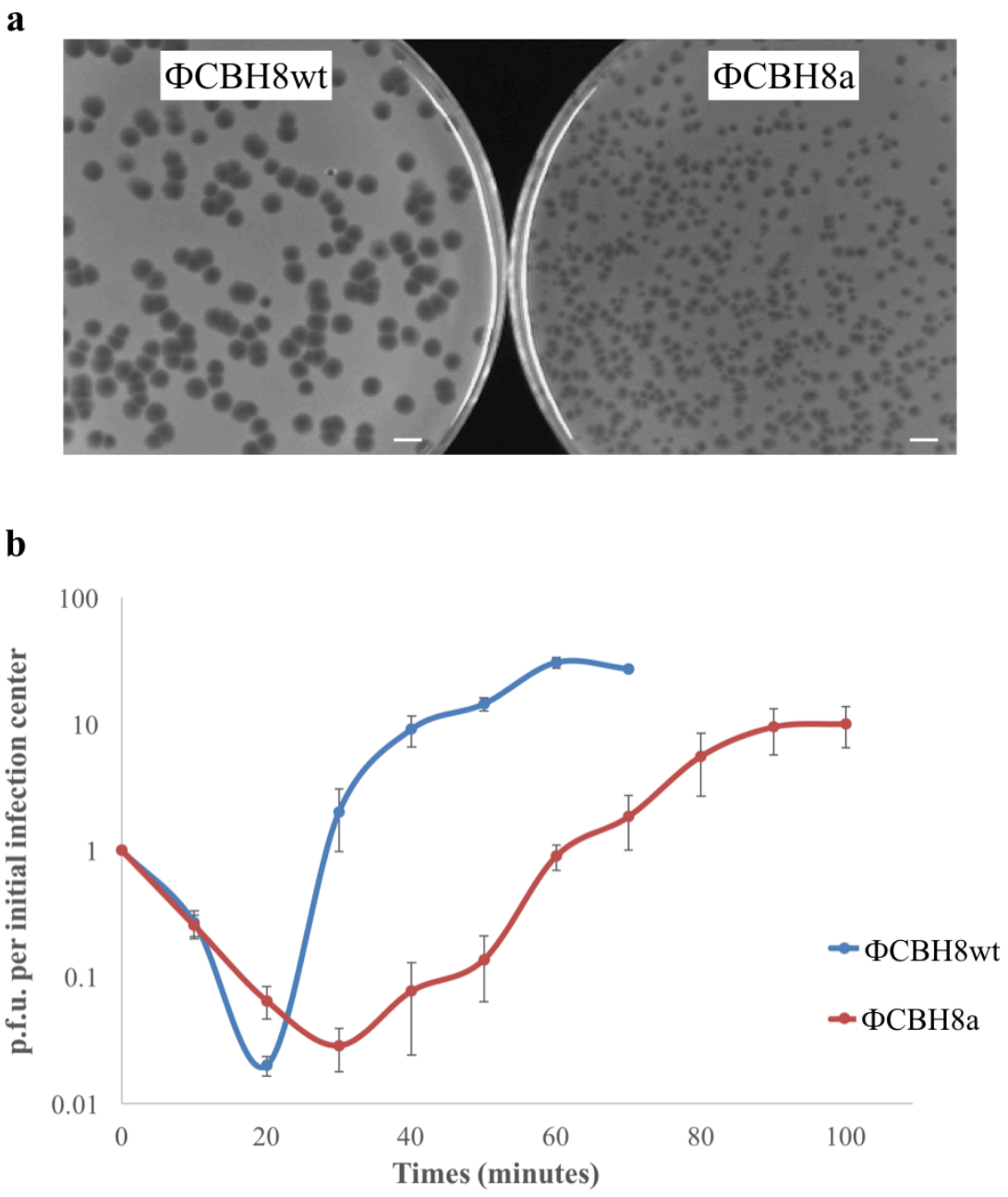


Figure 5.6: Fitness of AsiA mutant ΦCBH8a. (a) Plaque morphology comparison of ΦCBH8a and ΦCBH8wt grown on a top lawn of *S*39006. Scale bars represent 2 mm. (b) Burst size comparison of ΦCBH8a and ΦCBH8wt. Phages were added to exponential phase *S*39006 culture at a MOI of 0.1, and phage titer was measured every 10 minutes. Burst size was calculated by dividing the number of phages in the culture by the number of phages at time point 0. Details of the burst size assay is described in Section 3.10, Chapter 2. Error bars represent standard deviation from triplicates.

4. The effect of AsiA on ToxIN$_{Pa}$ promoter activity

Considering the effect of *asiA* mutation on both host growth and phage fitness, the above results suggest that the involvement of the AsiA protein in ToxIN$_{Pa}$-mediated Abi might be related to its function in σ-appropriation. Previous studies suggested that ToxIN$_{Pa}$ is a highly dynamic complex with newly transcribed antitoxin RNA constantly being cleaved by the toxin, thereby inhibiting ToxN$_{Pa}$ toxicity (91, 92). Therefore, perturbation of *toxIN$_{Pa}$* transcription could affect the continuous supply of ToxI$_{Pa}$, lead to the release of ToxN$_{Pa}$ and activation of ToxN$_{Pa}$-mediated Abi. Since AsiA in T4 was known to decrease host gene transcription and AsiA in ΦCBH8wt was shown to cause *S*39006 cell arrest, it is possible that AsiA could be involved in activating ToxN$_{Pa}$-mediated Abi by perturbing transcription from the ToxIN$_{Pa}$ promoter. To investigate this hypothesis, plasmid pTA104 (64) that contains the ToxIN$_{Pa}$ promoter (transcriptional positions -200 to +3)-*lacZ* fusion was first introduced into *E.coli* DH5α (Figure 5.7a). At the same time, pBAD33 (control), the wild type *asiA* cloned into pBAD33 (pBAD33::*asiA*-wt) or *asiA* from ΦCBH8a cloned into pBAD33 (pBAD33::*asiA*-CBH8a) was also introduced into *E.coli* DH5α carrying pTA104. *E.coli* DH5α, instead of *S*39006, was used in this experiment because of the difficulty in transforming *S*39006 with pTA104. The growth of each strain was assessed with *asiA* expression induced at 2h with 0.05% L-ara. As shown in Figure 5.7b, wild type *asiA* expression inhibited DH5α culture growth to an extent even greater than observed in *S*39006. Consistent with the results in *S*39006, expression of mutant *asiA* showed no inhibition of cell division and behaved like the control culture. To determine the level of ToxIN$_{Pa}$ promoter activity (reflected as β-galactosidase production), the β-galactosidase activity (normalized to cell density) in each culture was measured in a MUG assay. As shown in Figure 5.7c, β-galactosidase activity after induction of wild type AsiA was significantly lower compared to that of the control, with the greatest difference of 3-fold observed at 8h. In contrast, β-galactosidase activity of cells expressing the mutant AsiA was essentially identical to that of the control. This suggests that *asiA* expression could reduce ToxIN$_{Pa}$ promoter activity by as much as 3-fold in DH5α, whereas expression pf the mutant *asiA* had no such effect. These results are consistent with the hypothesis that the AsiA protein could trigger ToxIN$_{Pa}$-mediated Abi by perturbing *toxIN$_{Pa}$* transcription during σ-appropriation.

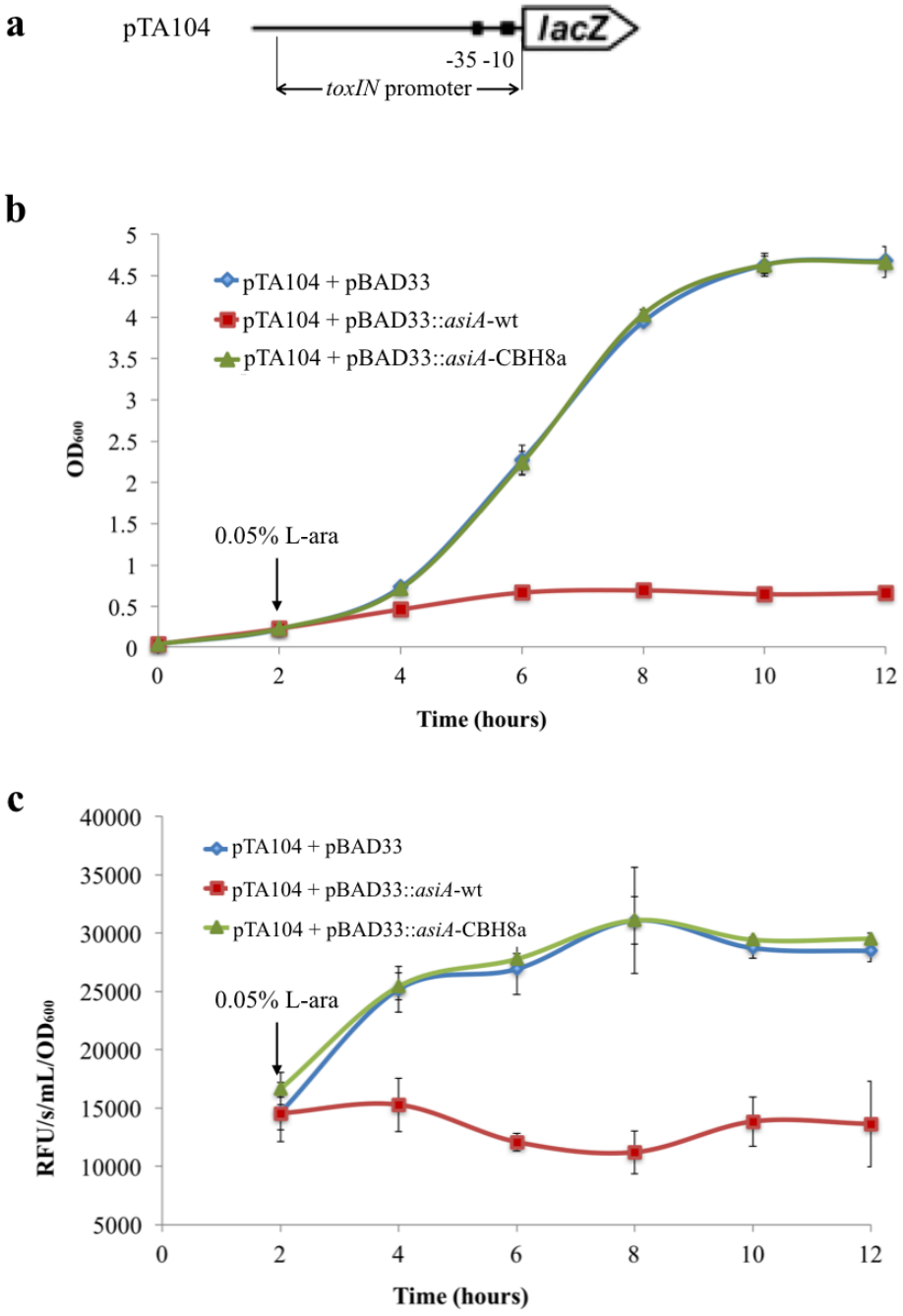


Figure 5.7: The effect of AsiA on ToxIN$_{Pa}$ promoter activity. (a) Schematic representation of the ToxIN$_{Pa}$ promoter-*lacZ* fusion in pTA104, using pRW50 as backbone (90). (b) Growth curve of *E.coli* DH5α transformed with pTA104 and pBAD33 (control), pBAD33::*asiA*-wt or pBAD33::*asiA*-CBH8a. *asiA* expression was induced at 2h with 0.05% L-ara and the OD_{600} of the cultures was measured every 2h until stationary phase. (c) MUG assay of cultures measuring β-

galactosidase activity. Detailed method can be found in Section 5.2, Chapter 2. In both (b) and (c), error bars represent standard deviation from triplicates.

5. Discussion

*S*39006 phages sensitive to ToxIN$_{Pa}$-mediated Abi could escape the system via mutation of the phage gene encoding the AsiA protein involved in σ-appropriation. Complementation of AsiA restored the Abi phenotype to escape phages, confirming that AsiA plays a role in phage sensitivity to ToxIN$_{Pa}$-mediated Abi. This is the first example where an *orf* associated with ToxIN$_{Pa}$ functionality has a defined function.

The majority of ΦCBH8 and ΦCHI14 mutants with an *asiA* mutation have the CTD of the protein extended or truncated while only one mutant had a NTD missense mutation. Both the NTD and CTD mutants lost their ability to inhibit host cell division. The results involving the V13G mutant are consistent with previous research showing that this residue is conserved in T4 and is important in σ-appropriation by contacting Region 4 of σ^{70} (162). In contrast, the results involving the CTD mutants of AsiA in this study differ from previous research. According to previous work on T4, the CTD of AsiA is not essential for its function because mutants with a 17-residue truncation still showed moderate level of toxicity, a 28-residue truncation still showed low level of toxicity and only when 34-residues in the CTD were truncated did AsiA show no toxicity at all (162). However, in the work described here, expression of *asiA* from ΦCBH8a with a 19-residue truncation at the CTD showed no toxicity to *S*39006, suggesting that the CTD residues are required for AsiA function. This could mean that the functionality of the AsiA CTD differs in different phages. Nevertheless, these results show that, at least in ΦCBH8, the CTD residues of AsiA are important for σ-appropriation. This view is further corroborated by results from the burst size assay where the CTD truncation of AsiA reduced phage burst size by 60% and extended the latent period. Similar fitness defect was also observed in T4 where a mutant with AsiA truncated after the 22nd residue (from the N-terminus) had a significantly longer latent period (163). Taken together, these results suggest that the mutated AsiA in ToxIN$_{Pa}$-escaping mutants of ΦCBH8 and ΦCHI14 have completely lost their ability to inhibit host cell growth, whereas they are still able to facilitate phage propagation, albeit with significantly decreased efficiency (as seen in the case of ΦCBH8a).

The compromised functionality of AsiA in ToxIN$_{Pa}$-escape mutants of ΦCBH8 and ΦCHI14 suggest that the mechanism for AsiA activation of ToxIN$_{Pa}$-mediated Abi may be closely related

to its function in σ-appropriation. Three possible mechanisms can be proposed: (1) AsiA perturbs *toxIN$_{Pa}$* transcription, causes imbalance of ToxN$_{Pa}$: ToxI$_{Pa}$ ratio which leads to ToxIN$_{Pa}$ disassembly; (2) The toxicity of AsiA in *S*39006 triggers some host response pathway that subsequently activates ToxIN$_{Pa}$-mediated Abi; (3) ToxIN$_{Pa}$-mediated Abi is directly activated by another phage middle gene product(s) but AsiA is essential for its transcription. This study has revealed evidence for the first mechanism because the effect of AsiA on ToxIN$_{Pa}$ promoter activity showed that wild type AsiA caused a significant reduction in ToxIN$_{Pa}$ promoter activity, whereas mutant AsiA had no such effect, suggesting that ΦCBH8 and ΦCHI14 mutants may escape ToxIN$_{Pa}$ because their AsiA no longer perturbs *toxIN$_{Pa}$* transcription. Future work could focus on directly measuring the ratio of ToxN$_{Pa}$ and ToxI$_{Pa}$ in the presence of AsiA, as well as testing Abi of previously insensitive *S*39006 phages in the presence of AsiA from ΦCBH8 or ΦCHI14. Mechanism 2 is also possible because of the significant host growth arrest caused by AsiA, similar to that observed for some LD *orf*s in Chapter 4. While mechanism 3 is also plausible, ΦCBH8a fitness assays suggest that the mutant phages could still produce progeny and it is therefore unlikely that transcription of the putative middle gene products are obliterated.

In summary, the discovery that AsiA is involved in activating ToxIN$_{Pa}$-mediated Abi shows that σ-appropriation, a mechanism benefiting phage proliferation, could also be involved in phage demise. This further complicates the story of whether activation of ToxIN$_{Pa}$-mediated Abi is caused by targeted phage product(s) *per se* or is just the indirect effect of normal phage processes, or perhaps both, depending on the specific phage and host.

Chapter 6 - Host dependency of ToxIN$_{Pa}$ sensitivity and characterization of ΦCBH8 receptor

1. Introduction

During the interaction between phages and Type III TA systems such as ToxIN$_{Pa}$, a phage product(s) might directly activate ToxIN$_{Pa}$-mediated Abi but might also indirectly activate Abi through triggering certain response pathway(s) in the host. If the latter turns out to be true, then ToxIN$_{Pa}$-sensitivity could be host dependent. However, no data supports a host-dependency of ToxIN$_{Pa}$-sensitivity so far. While it has been observed that it is relatively easier to isolate ToxIN$_{Pa}$-sensitive phages for some hosts (such as *P.atrosepticum* and *E.coli K-12*) (64) and difficult for others (such as *D. solani* MK10, where no ToxIN$_{Pa}$-sensitive phage has been isolated, so far), this is not proof that ToxIN$_{Pa}$-sensitivity is host dependent, because ToxIN$_{Pa}$ might not be adequately expressed and correctly assembled in some hosts (Xinzhe Fang, unpublished data). Therefore, it is important to test the same phage for ToxIN$_{Pa}$-sensitivity in different hosts where ToxIN$_{Pa}$ is known to be active for Abi.

The first step during phage infection is receptor recognition and adsorption. Phages have been found to utilize a wide range of cell surface components as receptors, including membrane proteins that function as porins or flagella that function as motility nanomachines (16). The host range of phages differ with some phages able to only infect very specific species of bacteria while others are able to infect a broader range of hosts. Both ΦM1 and ΦTE, the only well-studied ToxIN$_{Pa}$-sensitive phages so far, are very sub-species specific (164), therefore limited information is available on their susceptibility to ToxIN$_{Pa}$ in more distant hosts. *S*39006 phages isolated in this study were also found to have a narrow host range (Section 4, Chapter 3). Since receptor specificity is the first limiting step to successful infection (32), it may be possible to artificially expand the host range of certain phages by modifying their tail structure (165, 166). It is also possible to render a host susceptible to phage by artificially introducing the appropriate receptor.

The aim of this chapter is to identify the receptor(s) used by *S*39006 phages and express the receptor in a different host to try to expand the host range of these phages. This would enable us to test the hypothesis of whether Abi by ToxIN$_{Pa}$ depends on the host background, and whether relevant ΦCBH8 loci involved in Abi-activation in a new host is the same as in *S*39006. Expansion of the host range of *S*39006 phages would also enable exploitation of these phages as biotechnological tools if they have the capacity to transfer genetic material between bacteria. Finally, exploration of the native function of a *S*39006 phage receptor and its regulation will not only deepen insight into the function of bacterial surface components but also provide novel ways to manually control host susceptibility to phage, with potential biotechnological value.

2. ΦCBH8 utilizes OmpW as *S*39006 receptor

In order to identify the receptor that ΦCBH8 utilizes to infect *S*39006, ΦCBH8-resistant mutants of *S*39006 were generated and the mutated gene(s) was identified as a potential receptor gene(s). Mutagenesis using the transposon Tn-KRCPN1 was performed to generate *S*39006 mutants, as described in Section 4.2.5, Chapter 2 (33). Theoretically, Tn-KRCPN1 would randomly insert into the *S*39006 genome and mutants that had successfully acquired the transposon would contain kanamycin resistance. If Tn-KRCPN1 inserted into the phage receptor, then mutants would also be resistant to ΦCBH8. Therefore, in order to select for mutants with Tn-KRCPN1 insertion in receptor gene(s), cells subjected to mutagenesis were grown in the presence of kanamycin and high titer ΦCBH8 in a top lawn.

After 48h incubation, four mutant colonies arose, three of which were found to be false positives after re-streaking and testing for ΦCBH8 susceptibility. One mutant, named *S*39006-R1, was truly resistant to ΦCBH8 as well as possessing kanamycin resistance (Figure 6.1a-b). An adsorption assay was carried out to compare ΦCBH8 adsorption to wild type *S*39006 (*S*39006-wt) and *S*39006-R1. As shown in Figure 6.2, at a MOI of 0.01, more than 90% of ΦCBH8 have adsorbed to *S*39006-wt within 20min (n=3), as reflected in the decreased abundance of free phage remaining in the culture. In contrast, no phage has adsorbed to *S*39006-R1, as in the phage-only control. This suggests that the mutation in *S*39006-R1 has made the bacteria resistant to ΦCBH8 by preventing adsorption.

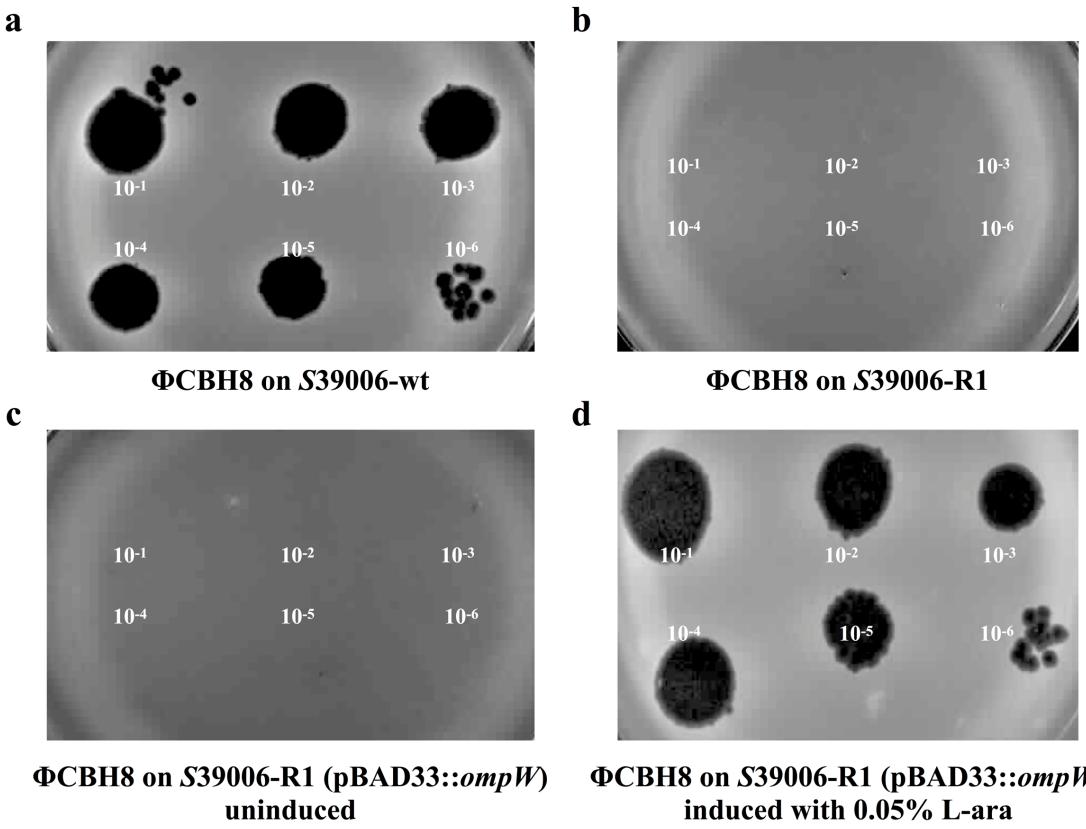


Figure 6.1: Spot test assay of ΦCBH8 infecting *S*39006-R1 with and without complementation of the putative receptor mutant. (a) ΦCBH8 forms clear spots on *S*39006-wt; the positive control. (b) ΦCBH8 does not form clear spots on *S*39006-R1; indicating resistance. (c) ΦCBH8 does not form clear spots on *S*39006-R1 carrying the *ompW* plasmid but without induction of expression**.** (d) ΦCBH8 forms clear spots on *S*39006-R1 complemented with *ompW* with induction, showing restoration of susceptibility

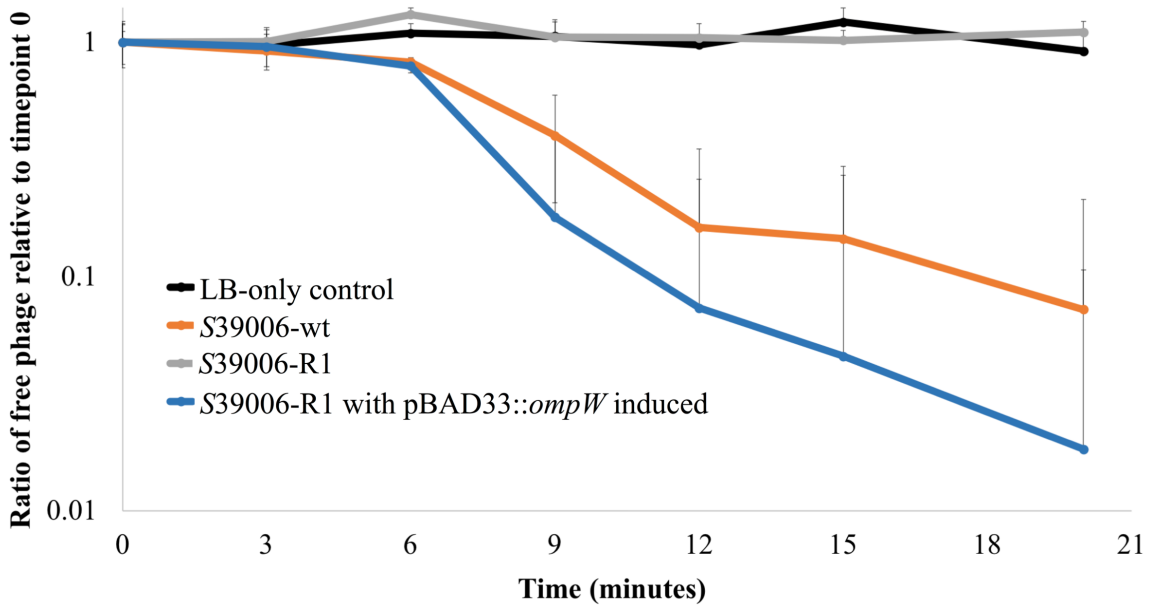


Figure 6.2: Adsorption assay of ΦCBH8 infecting *S*39006-R1 with and without complementation of the putative receptor mutant. The adsorption assay was carried out as described in detail in Section 3.9, Chapter 2. The lower the amount of free phage in the culture, the higher the number of phages that have adsorbed to cells. The percentage of free phage is calculated by dividing the titer of the phage taken from the sample at any given time point by the titer of the phage in the sample at time point 0. Error bars represent standard deviation from triplicates.

Replicon cloning (33) was used to investigate which gene was disrupted by the transposon, details of the method is described in Section 4.2.6, Chapter 2. Genomic DNA from *S*39006-R1 was extracted and digested with the restriction enzyme EcoRV which does not cut within the transposon, followed by ligation which allowed the digested DNA fragments containing the transposon to circularize. Because Tn-KRCPN1 contains an origin of replication (*oriR6K*), the circularized DNA (replicon) could be propagated within *pir*+ strains and this allows for cloning, sequencing, and then identification of *S*39006 sequences adjacent to the transposon insertion site. Sequencing results showed that Tn-KRCPN1 was inserted into the *ompW* gene that encodes an outer membrane protein W (OmpW) containing 211 amino acids. Insertion of the transposon after the 155th amino acid disrupted translation of *ompW*, therefore preventing phage adsorption. This suggests that OmpW is the host membrane component recognized by ΦCBH8 during adsorption.

To confirm that OmpW is the phage receptor, the *ompW* gene from *S*39006-wt (denoted *ompW*$_{S39006}$) was first amplified by PCR and inserted into the multi-cloning site of an L-ara inducible pBAD33 plasmid. The recombinant plasmid was introduced into *S*39006-R1 by transformation and its infection by ΦCBH8 was assessed in a spot test. As shown in Figure 6.1c-d, induction of *ompW* expression from the pBAD33::*ompW*$_{S39006}$ plasmid with 0.05% L-ara restored *S*39006-R1's susceptibility to ΦCBH8. Similarly, ΦCBH8 adsorption to *S*39006-R1 (pBAD33::*ompW*$_{S39006}$) was also restored to a level comparable to *S*39006-wt (Figure 6.2). The complementation result confirmed that OmpW is the receptor for ΦCBH8.

Spot tests and adsorption assays with *S*39006-wt, *S*39006-R1 and *S*39006-R1 (pBAD33::*ompW*$_{S39006}$) were also performed using ΦCHI14, ΦX20 and ΦCBH189. Similar to ΦCBH8, the other *S*39006 phages also failed to lyse or adsorb to OmpW-deficient mutant whereas complementation of *ompW* restored their ability to adsorb and infect (data not shown). This shows that the ToxIN$_{Pa}$-sensitive *S*39006 phages characterized so far all utilize OmpW as the receptor in *S*39006.

3. Expansion of ΦCBH8 host range

BLASTp search for homologues of *S*39006 OmpW showed that OmpW is conserved among Gram-negative species. In particular, OmpW from several *Dickeya, Yersinia* and *Pectobacterium* species share high protein sequence identity (>70% identity with 100% coverage) with

OmpW$_{S39006}$. Protein sequence alignment of OmpW$_{S39006}$ with OmpW from some of these strains showed that conserved residues and residues with highly similar properties are dispersed throughout the protein sequence (Figure 6.3). Attempts were made to transform some of these strains and other commonly used laboratory strains (e.g. *E.coli* DH5α) with pBAD33::*ompW*$_{S39006}$ to test whether ΦCBH8 could then infect these otherwise non-susceptible hosts. As shown in Table 6.1 and Figure 6.4, of the 16 strains tested, *E.coli* DH5α, *D.solani* MK10 and *D.zeae* NCPP 3532 became susceptible to ΦCBH8 after *ompW*$_{S39006}$ expression was induced. The role of OmpW as the phage receptor in these strains was further confirmed in an adsorption assay. As shown in Figure 6.5, at a M.O.I. of 0.01, within 20 min more than 98% of ΦCBH8 adsorbed to *E.coli* DH5α after induced expression of *ompW*$_{S39006}$, comparable to ΦCBH8 adsorption to *S*39006-wt (Figure 6.2). Similarly, in the presence OmpW$_{S39006}$, ΦCBH8 adsorption to *D.zeae* NCPP 3532 was also rapid, with more than 94% of phages adsorbed within 20 min. In contrast, ΦCBH8 adsorption to *D.solani* MK10 expressing *ompW*$_{S39006}$ was slower, with 80% of phages adsorbed within 20 min and 88% adsorbed after 25 min. The adsorption results provide important confirmation that expression of *ompW*$_{S39006}$ in previously resistant bacterial strains could artificially expand the host range of ΦCBH8 by providing the appropriate receptor absent from those hosts. This shows that the only obstacle preventing ΦCBH8 from successfully infecting those alternative hosts is the lack of receptor, instead of the presence of other post-adsorption defense mechanisms.

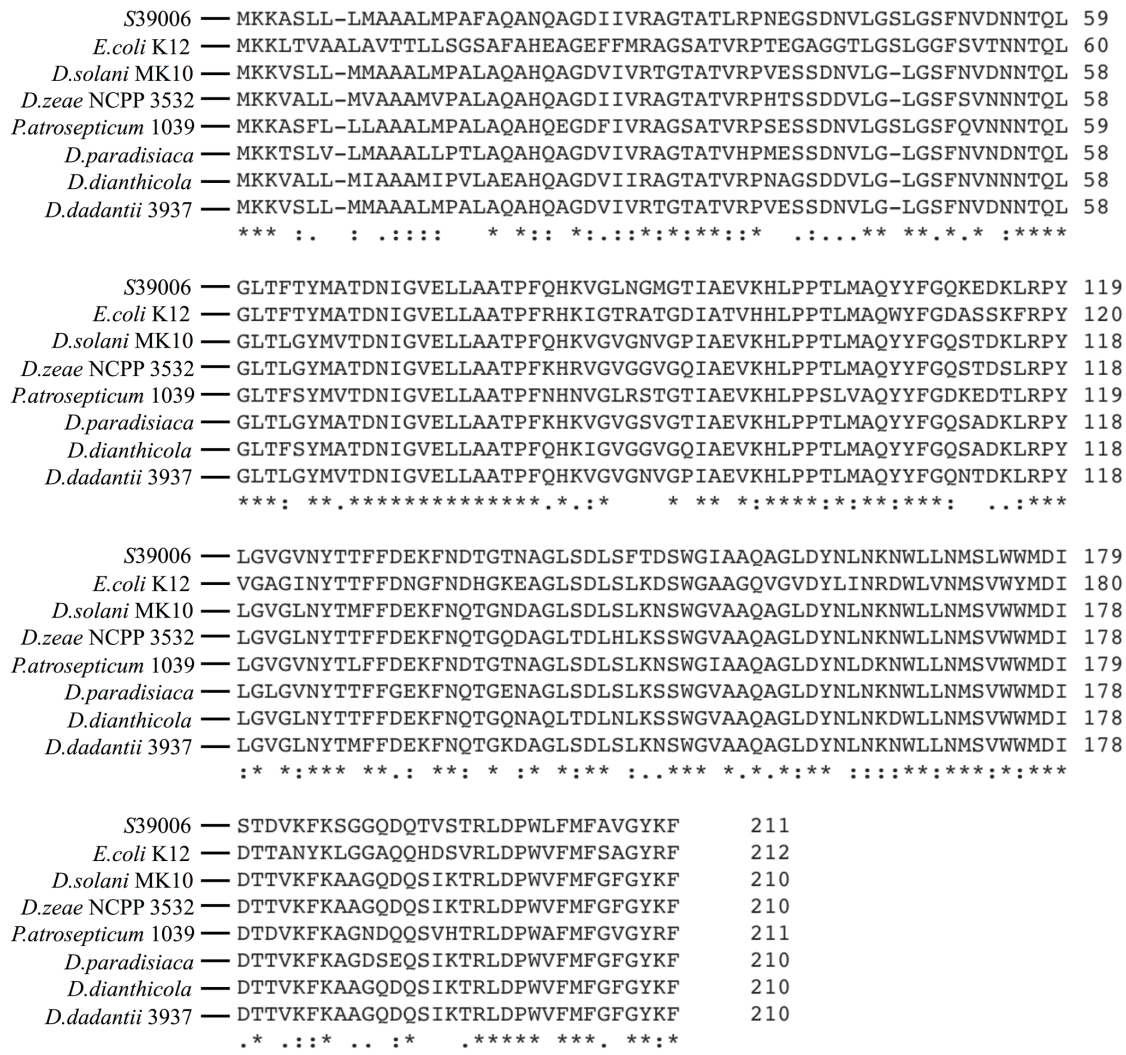


Figure 6.3: Protein sequence alignment of OmpW from *S*39006 and other strains. "*" (asterisk): positions that have a single, fully conserved residue. ":" (colon): conservation between groups of strongly similar properties. "." (period): conservation between groups of weakly similar properties.

Table 6.1: Transformation of naturally resistant hosts with the recombinant plasmid pBAD33::*ompW*$_{S39006}$.

Strains tested	Transformable with pBAD33::*ompW*$_{S39006}$?*	Sensitive to ΦCBH8?
Pantoea agglomerans 9rz4	✓	×
Serratia plymuthica A153	✓	×
Serratia marcescens Sma12	✓	×
Dickeya sp. NCPPB 3274	✓	×
Dickeya paradisiaca	✓	×
Dickeya solani MK10	✓	✓
Pectobacterium atrosepticum SCRI1043	✓	×
E. coli DH5α	✓	✓
Dickeya zeae NCPP 3532	✓	✓
Dickeya dadantii 3937	×	×
Pectobacterium carotovorum sp. ATCC39048	×	×
Dickeya dianthicola	×	×
Dickeya chrysanthemi 402	✓	×
Citrobacter rodentium DBS100	✓	×
Pectobacterium carotovorum SCRI 193	✓	×
Pseudomonas aeruginosa PAO1	✓	×

**Note: strains transformed with pBAD33::ompW$_{S39006}$ were subjected to plasmid mini-prep and subsequent sequencing to ensure that pBAD33::ompW$_{S39006}$ was obtained and that ompW$_{S39006}$ did not mutate. In the absence of the appropriate antibody, it is unknown whether OmpW$_{S39006}$ is expressed in each strain, therefore only the ones that became sensitive to ΦCBH8 were chosen for further characterization.*

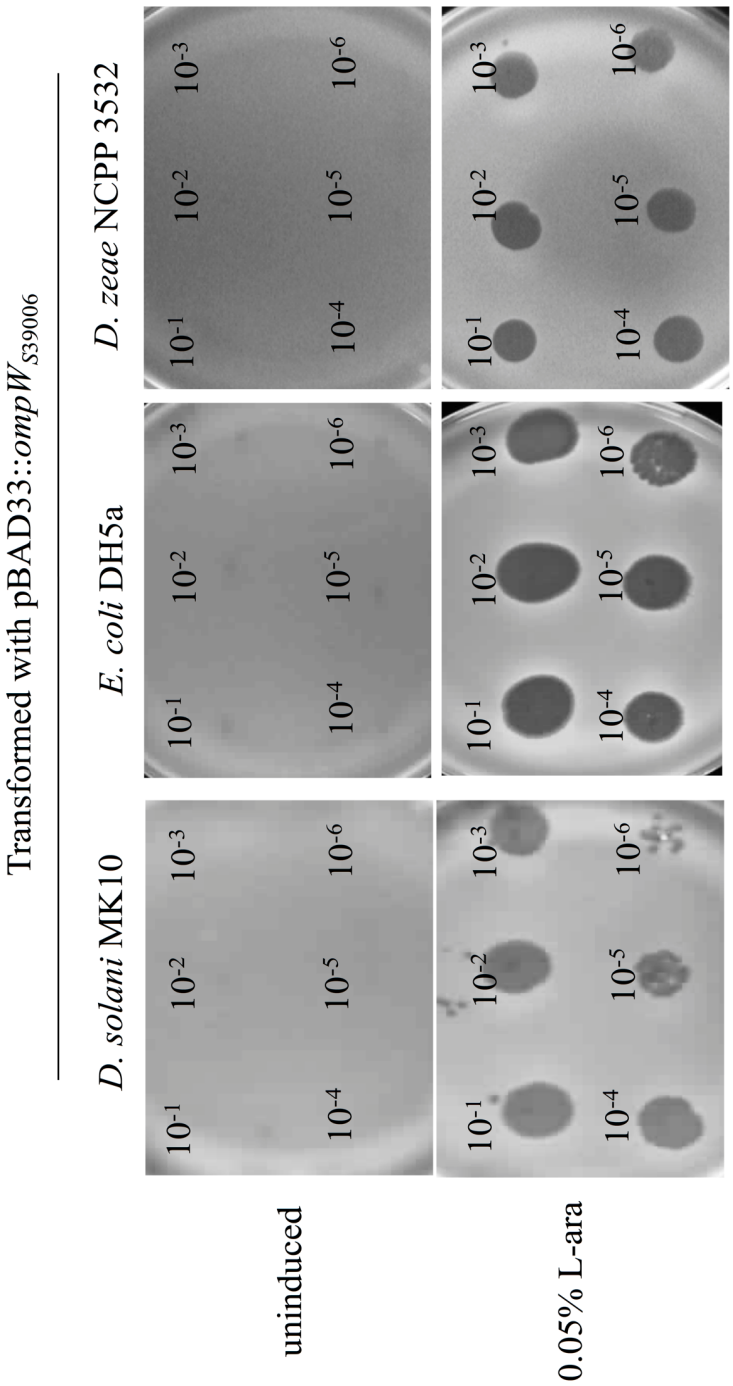


Figure 6.4: Spot test of serially diluted ΦCBH8 infecting *E.coli* DH5α, *D.solani* MK10 and *D.zeae* NCPP 3532 transformed with pBAD33::*ompW*$_{S39006}$, with and without L-arabinose induction. Without induction of *ompW*$_{S39006}$ expression, the hosts remained resistant to ΦCBH8 (top row). However, when induced with 0.05% L-ara, ΦCBH8 was able to form clear spots on the hosts (bottom row).

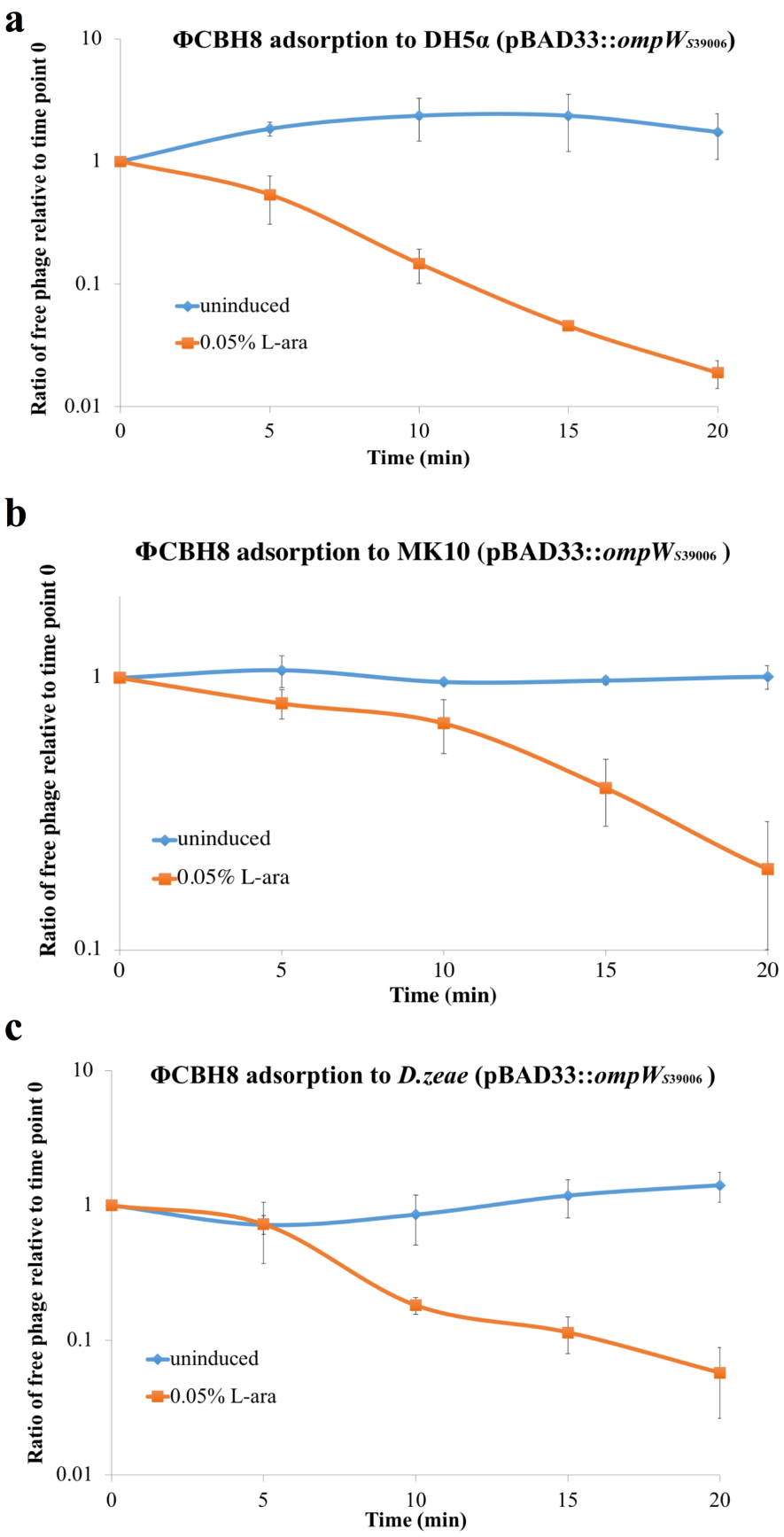


Figure 6.5: (figure legend provided overleaf)

Figure 6.5: Adsorption of ΦCBH8 to (a) *E.coli* DH5α, (b) *D.solani* MK10 and (c) *D.zeae* NCPP 3532 transformed with pBAD33::*ompW*$_{S39006}$, with and without L-ara induction. The adsorption assay was carried out as described in detail in Section 3.9, Chapter 2. The lower the amount of free phage in the culture, the higher the number of phages that have adsorbed to cells. The percentage of free phage is calculated by dividing the titer of the phage taken from the sample at any given time point by the titer of the phage in the sample at time point 0. Error bars represent standard deviation from triplicates.

4. ΦCBH8 does not activate ToxIN$_{Pa}$-mediated Abi in *E.coli* DH5α

With the expanded host range, it was possible to explore ToxIN$_{Pa}$-mediated Abi activation by ΦCBH8 in bacteria other than *S*39006. Previous studies showed that ToxIN$_{Pa}$ encoded by pTA46 is functional as an Abi system in *E.coli* DH5α (64), therefore DH5α was transformed with both pTA46 and pBAD33::*ompW*$_{S39006}$. As control, DH5α was transformed with both pTA47 (ToxIN$_{Pa}$-FS) and pBAD33::*ompW*$_{S39006}$. The EOP of ΦCBH8 infecting DH5α containing ToxIN$_{Pa}$ or ToxIN$_{Pa}$-FS was measured. With an EOP of 0.148 (n=3), it seemed that ΦCBH8 was not aborted by ToxIN$_{Pa}$. Phage plaques that arose from the top lawn of DH5α containing ToxIN$_{Pa}$ were bulked up and plated out again on DH5α containing ToxIN$_{Pa}$ or ToxIN$_{Pa}$-FS to rule out the possibility that they were high frequency escape mutants. Indeed, the EOP of escape "suspects" was still 10^{-1}, indicating that they were not escape mutants. Therefore ΦCBH8 does not activate Abi in *E.coli* DH5α, suggesting that ToxIN$_{Pa}$-sensitivity is host dependent for this phage.

5. ΦCBH8 mediates plasmid transfer between hosts through transduction

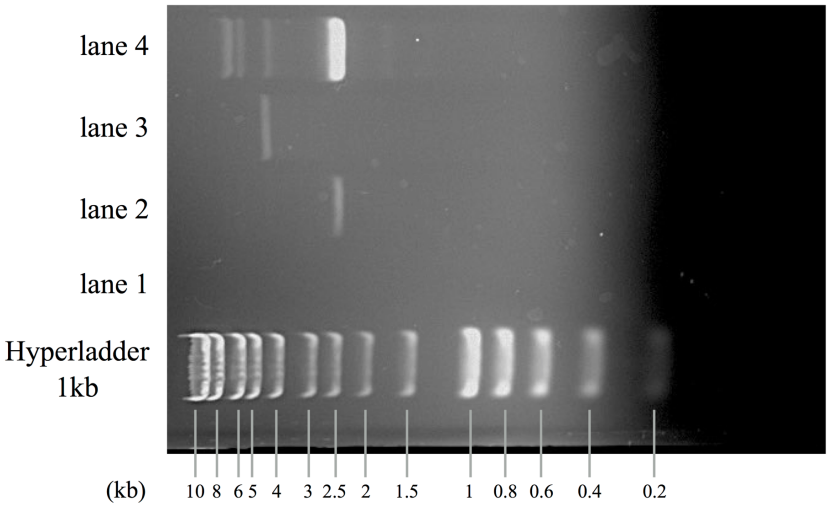
Some phages mediate the transfer of genetic material between bacteria through transduction and have been used as important biotechnological tools (Section 1.3.3, Chapter 1). Given the expanded host range of ΦCBH8, it was possible to test whether ΦCBH8 is a transducing phage in its native host and in new hosts. A transduction assay (described in detail in Section 3.8, Chapter 2) was first done to test whether ΦCBH8 could transduce the pBR322 plasmid (Ap^R, Tc^R) in *S*39006-wt. The donor strain was *S*39006-wt transformed with pBR322 and the recipient strain was *S*39006-wt without any plasmid. ΦOT8, a known high-efficiency transducer of *S*39006 was used as positive control and the recipient strain added with only phage buffer was used as negative control. Successful transductants were recognized by the ability to grow on LBA containing both ampicillin and tetracycline. The result from the transduction assay showed that ΦCBH8 is able to transduce pBR322 between *S*39006-wt, with the highest transduction efficiency of 4.4×10^{-8} (n=3) achieved at a MOI of 0.1. Next, the transduction of pBR322 in DH5α was tested. The donor strain was DH5α expressing *ompW*$_{S39006}$ and containing the pBR322 plasmid. The recipient strain was DH5α expressing *ompW*$_{S39006}$ but without any plasmid. DH5α containing the pBAD33::*ompW*$_{S39006}$ plasmid but without induction of expression was also used as a negative control recipient. Successful transductants were recognized by the ability to grow on LBA containing ampicillin, chloramphenicol and tetracycline. Results showed that ΦCBH8 is able to transduce pBR322 into DH5α, with the highest transduction efficiency of 5.0×10^{-9} (n=3) achieved at a MOI of 1. In all of

the transduction assays mentioned above, the receipt of the transduced plasmid was not only confirmed by the newly acquired antibiotic resistance of recipient cells, but also by the physical evidence of the acquired DNA visualized on agarose gel (Figure 6.6). For transduction of pBR322 into *S*39006-wt, plasmid mini-prep of transductant followed by digestion with HindIII yielded a single DNA band of ~5-6kb in size, confirming the acquisition of pBR322. For transduction of pBR322 into DH5α (expressing *ompW*$_{S39006}$), plasmid mini-prep of transductant followed by digestion with HindIII yielded two distinct DNA bands, one with ~5-6kb in size and the other with ~4-5kb in size, indicative of the existence of both pBR322 and pBAD33::*ompW*$_{S39006}$ in the recipient strain. To rule out the possibility of natural competence causing the acquisition of pBR322, additional negative controls were also used where phage lysates were treated with DNase and RNase to digest any genetic material in solution, as well as adding pBR322 DNA directly into the recipient cell culture. However, none of the negative controls yielded any positive results, proving that transduction by ΦCBH8 is indeed the reason for the observed plasmid transfer.

To further substantiate these results, a second plasmid (pTRB30, Kn^R, modified from pQE80L) was used to test ΦCBH8 transduction. Similar to the case of pBR322, ΦCBH8 was also shown to mediate pTRB30 transfer in *S*39006-wt and in DH5α (expressing OmpW$_{S39006}$).

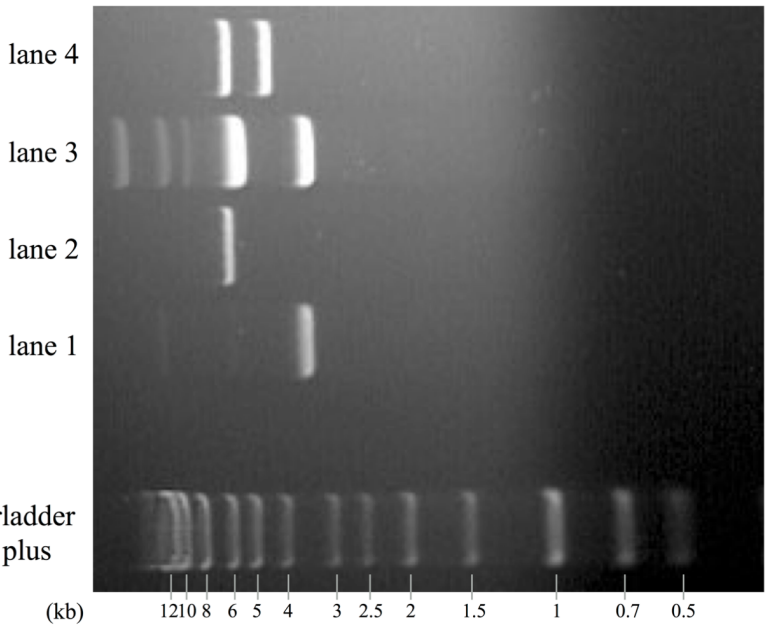
In contrast, ΦCBH8 was not able to transduce chromosomal markers. For both markers tested (Tn-KRCPN1, Kn^R and mini-Tn5, Sm/Sp^R), no successful transductant was found while ΦOT8, the positive control, was able to transduce both chromosomal markers with high efficiency. These results indicate that ΦCBH8 is a transducing phage that is able to mediate plasmid transfer in different species.

a



lane 1: (control) mini prep of *S*39006-wt without any plasmid (recipient)
lane 2: (control) mini prep of *S*39006-wt with pBR322 (donor)
lane 3: mini prep of successful transductants, digested with HindIII
lane 4: mini prep of successful transductants, undigested

b



lane 1: (control) mini prep of DH5α containing pBAD33::*ompW*$_{S39006}$ (recipient), undigested
lane 2: (control) mini prep of DH5α containing pBAD33::*ompW*$_{S39006}$ (recipient), digested with HindIII
lane 3: mini prep of successful transductants, undigested
lane 4: mini prep of successful transductants, digested with HindIII

Figure 6.6: Visualization of DNA transfer in (a) *S*39006-wt and (b) DH5α. In (a), the transduction of pBR322 between wild type *S*39006 strains was tested. Transductant that arose in

the selection plate was re-streaked at least twice and subjected to mini-prep. The resulting DNA was visualized on an agarose gel (lane 4) and exhibited the same pattern as the mini-prep product from the donor strain (lane 2). The mini-prep product from transductant was also digested with *Hin*d III and showed a size between 4-5 kb on the agarose gel, consistent with the 4.3 kb size of pBR322. In (b), the transduction of pBR322 between DH5α was tested. Both the donor and recipient DH5α already contained the recombinant vector pBAD33::*ompW*$_{S39006}$. Transductant that arose in the selection plate was re-streaked at least twice and subjected to mini-prep. The resulting DNA was visualized on the agarose gel (lane 3). The resulting DNA was also digested with *Hin*d III and contained two bands with distinct sizes, one between 4-5 kb (consistent with the 4.3 kb size of pBR322) and another between 5-6 kb, consistent with the 5.4 kb size of pBAD33.

6. The function of OmpW$_{S39006}$

As a highly conserved minor porin, the function of OmpW has only been elucidated in *E.coli* K-12 as the receptor for colicin S4 (167). OmpW has also been connected with resistance to antibiotics, resistance to phagocytosis and the transition from aerobic to anaerobic growth conditions (168-170). However, the exact role of OmpW in relation to these phenotypes is unclear. The function of OmpW in *S*39006 has not been studied before, therefore it would be important to gain some insight into the role of OmpW other than as a phage receptor. This would enable us to exploit OmpW functionality to regulate phage infection.

6.1 OmpW$_{S39006}$ is not important for *S*39006 fitness and is non-toxic

The viability of *S*39006-R1 (with transposon inserted into *ompW*) serves as proof that OmpW is not essential for *S*39006 survival. To further assess whether OmpW is involved in bacterial growth and affects bacterial fitness, a growth assay was carried out in LB medium comparing wild type *S*39006 versus an OmpW mutant *S*39006-R4, where 4/5 of the OmpW peptide has been truncated and therefore should be non-functional. Detailed method on the genesis of *S*39006-R4 is in Section 7.1, Chapter 2. As shown in Figure 6.7a-b, no difference in growth has been observed under aerobic or anaerobic conditions between *S*39006-wt and *S*39006-R4 grown in LB. Phase contrast microscopy imaging of *S*39006-wt and *S*39006-R4 also showed no difference in cell morphology during growth (data not shown). This shows that OmpW does not affect cell growth and division under either aerobic or anaerobic conditions with rich medium.

The toxicity of OmpW$_{S39006}$ in a heterologous host was also investigated. In a growth assay of *E.coli* DH5α, cells containing pBAD33::*ompW*$_{S39006}$ were induced to express OmpW after 2 hours of growth with 0.5% L-ara. As shown in Figure 6.7c, cells with induced OmpW expression exhibited a similar growth rate as cells containing the empty pBAD33 vector. In contrast, DH5α containing the empty pBAD33 vector but without L-ara induction reached a higher OD_{600}, due to the toxic effect of L-ara on cells. Taken together, these results indicate that OmpW is not toxic even when over-expressed in *E.coli* DH5α.

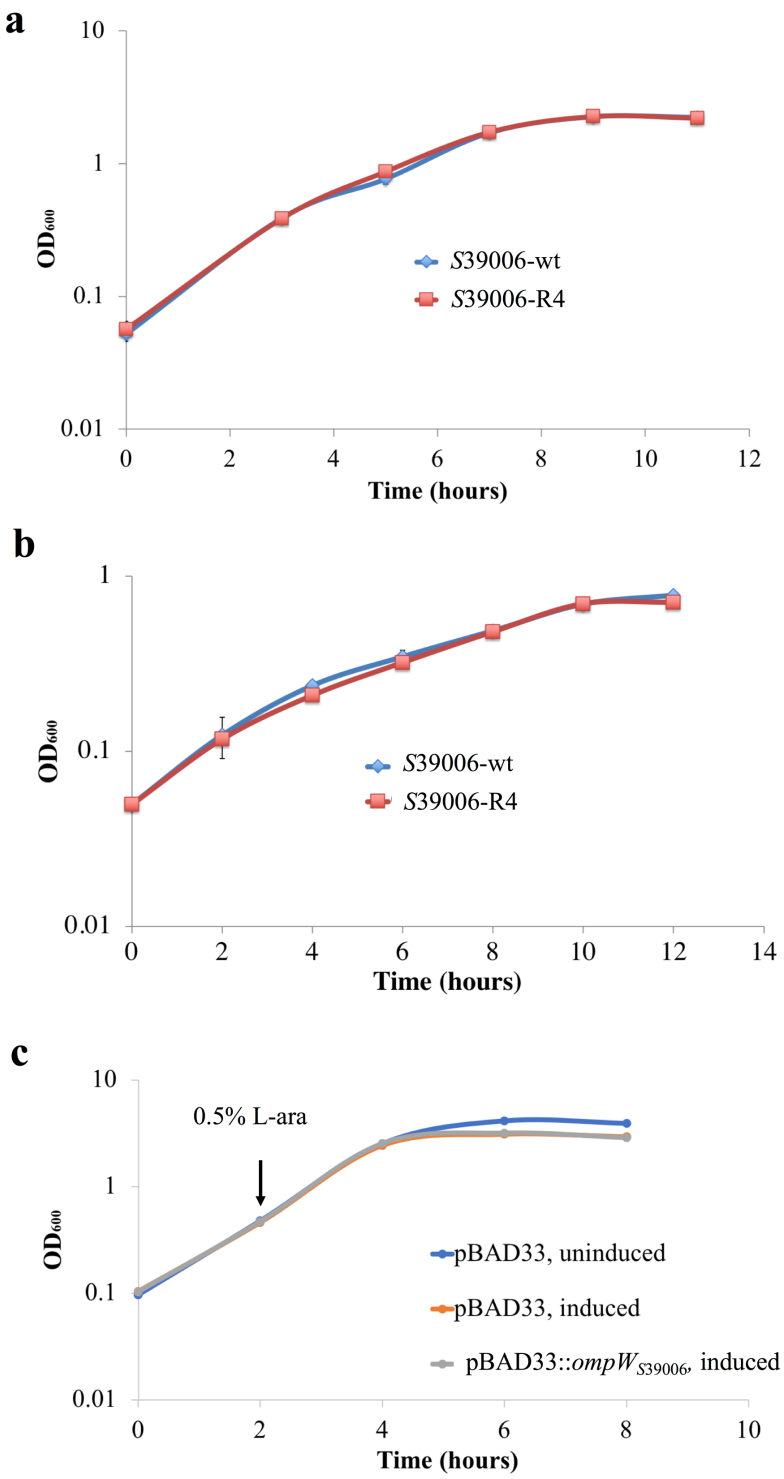


Figure 6.7: Growth assay of cells without functional OmpW$_{S39006}$ or over-expressing OmpW$_{S39006}$. (a) Growth curve of *S*39006-wt versus *S*39006-R4 in aerobic conditions, with a starting OD of 0.05. (b) Growth curve of *S*39006-wt versus *S*39006-R4 in anaerobic conditions, with a starting OD of 0.05. (c) Growth curve of DH5α containing pBAD33 or pBAD33::*ompW*$_{S39006}$, with or without 0.5% L-ara induction, with a starting OD of 0.1.

6.2 OmpW is not associated with common phenotypes of *S*39006

*S*39006 is known for its phenotypes such as secretion of secondary metabolites, motility, and virulence. To investigate whether OmpW is involved in any of those phenotypes, a series of phenotypic assays were carried out to compare *S*39006-wt with *S*39006-R4, including secretion of Carbapenem, prodigiosin and N-butanoyl-L-homoserine lactone (BHL); ability to swim and swarm in agar; response to stress conditions such as the presence of ethanol, H_2O_2, SDS and NaCl; resistance to antibiotics; and virulence in potato and *C. elegans*. However, *S*39006-wt and *S*39006-R4 showed no significant difference in any of the phenotypes tested, suggesting that OmpW is not involved in secondary metabolite secretion, motility, stress tolerance and virulence phenotypes mentioned above. Detailed description and results of the phenotypic assays are included in Appendix III.

6.3 Genetic context and promoter analysis of *ompW*$_{S39006}$

In the absence of phenotypic differences between *S*39006-wt and *S*39006-R4, the genetic context of *ompW*$_{S39006}$ was analyzed to gain more insight on its function. As shown in the schematic representation of *ompW* and neighboring genes in Figure 6.8a, *ompW* is not part of an operon. Genes located upstream of *ompW* (gene encoding cysteine cluster protein, gene encoding a hypothetical protein, the *ycib* gene that encodes intracellular separation protein A and gene encoding acetyl-CoA esterase) are also not part of an operon, nor are they obviously related to each other functionally. Genes located downstream of *ompW* are part of the tryptophan operon and encodes for tryptophan synthetase. However, it is unlikely that *ompW* is related to the Trp operon because of the distance between them and the different direction of transcription.

In addition, sequences surrounding the promoter region of *ompW*$_{S39006}$ are were analyzed by BPROM to look for potential binding sites for regulators. As shown in Figure 6.8b, the promoter region contains two near-perfect binding sites for FNR (Fumarate and Nitrate Reductase Regulatory) protein, which is required for the switch from aerobic to anaerobic metabolism. This is consistent with results by Xiao and colleagues showing that OmpW is involved in aerobic to anaerobic transition in *E.coli* K-12 (169). However, previous results (Figure 6.7a-b) did not support the view that OmpW plays a role in bacterial growth in different oxygen availability

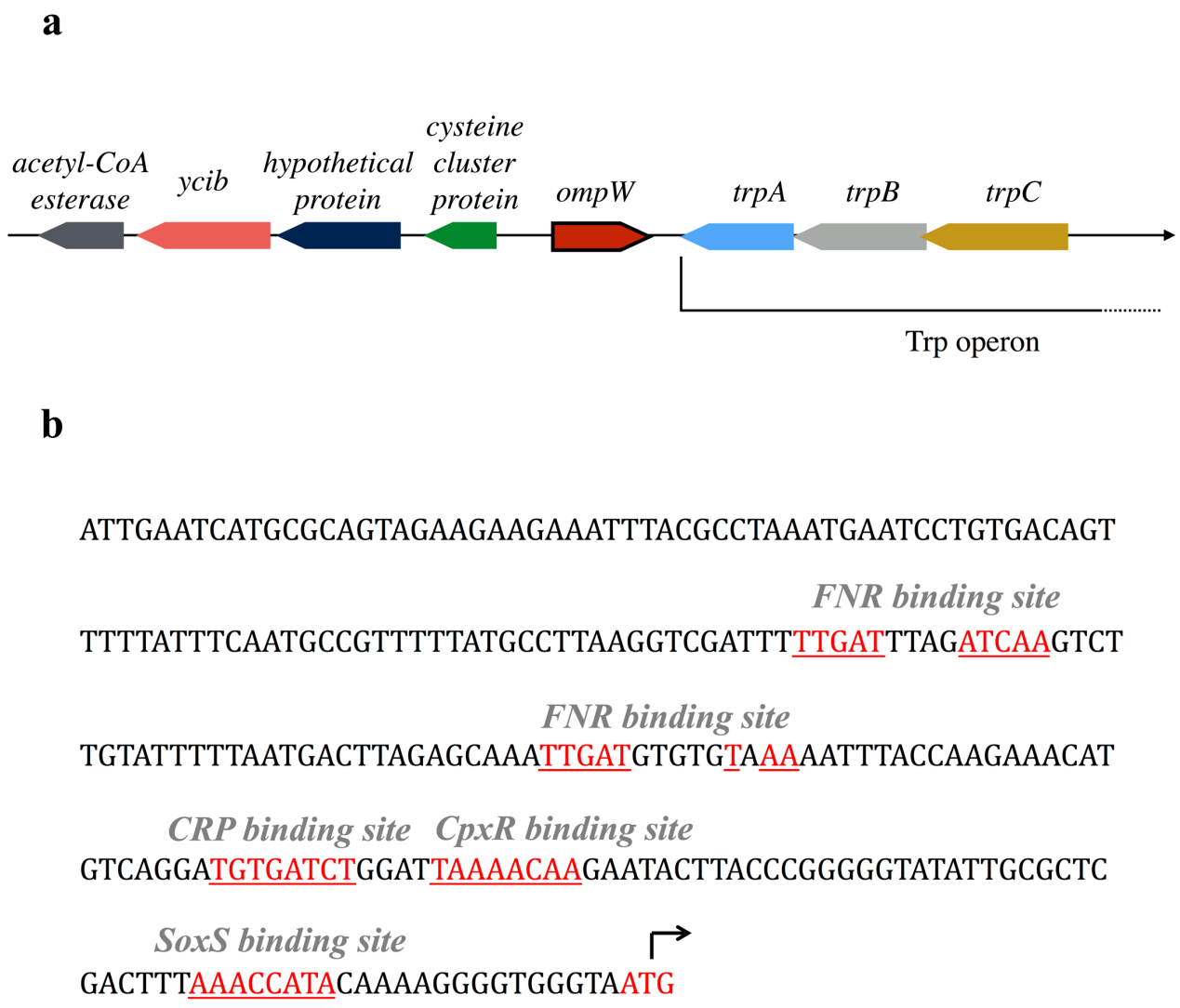


Figure 6.8: Genetic context and promoter analysis of *ompW*$_{S39006}$. (a) Schematic diagram showing genes neighboring *ompW* in *S*39006. The direction of arrow head represents 5'→3' direction of genes. (b) Putative regulator binding sites of *ompW* promoter region. Sequences in red color represent cognate consensus sequences for regulator binding. Arrow marks the presumed *ompW* transcription start site.

in *S*39006, at least in rich medium. The *ompW*$_{S39006}$ promoter region also contains putative binding sites for CpxR and SoxS proteins that activate stress-combative genes and superoxide response regulon, respectively. However, previous results (Figure S2, Appendix III) also provide contradictory evidence as the OmpW-deficient mutant failed to exhibit any growth difference in the presence of various stress conditions. Finally, a predicted CRP (cAMP receptor protein) binding site is also found 5' prior to the *ompW* transcription start site. Genes regulated by CRP are mostly involved in energy metabolism, such as the galactose, citrate, or PEP group translocation system (171), therefore OmpW could be involved in energy metabolism as a transporter. However due to the vast number of genes CRP regulates, it is not immediately clear what role OmpW might play.

6.4 Conservation of residues important for OmpW function

The structure of OmpW$_{S39006}$ was investigated next in order to understand its function. Attempts were made to impose the tertiary structure of OmpW$_{S39006}$ modelled by i-TASSER with structures of OmpW in *E.coli* K-12 and *P. aeruginosa* PAO1, the only two organisms whose OmpW structure has been elucidated by crystallography or NMR (172-174). However, such comparison has little meaning since i-TASSER relied on K-12 and PAO1 OmpWs to model the structure of OmpW$_{S39006}$, therefore potential differences could have been omitted. In contrast, comparison of the primary sequence of OmpW in the said organisms could provide more insight. As shown in Figure 6.9a, OmpW protein sequences are extremely similar between the three species, with the exception of a few short stretches of dissimilar sequences that are mainly located in extracellular loop regions. Among the conserved residues, those that are found to embellish the hydrophobic lumen of OmpW in K-12 and PAO1 remained identical or similar in hydrophobicity at cognate positions in *S*39006 (Figure 6.9b, red arrows). The hydrophobic lumen in K-12 and PAO1 were proposed to act as the channel for small hydrophobic molecules to pass through (172, 173), therefore the conservation of the hydrophobic lumen in *S*39006 implies a similar function. In addition, OmpW in PAO1 is believed to contain a Proline that forms a lateral opening in the transmembrane region of OmpW where the transported molecule would exit (173). Analysis of protein sequence alignment shows that this Proline (Figure 6.9b, green arrow) is conserved in *S*39006 as well, suggesting the existence of a lateral opening. Taken together, these similarities suggest that akin to the predicted function of OmpW$_{K-12}$ and OmpW$_{PAO1}$, OmpW$_{S39006}$ could also function as a tunnel through which a small hydrophobic molecule passes through the outer membrane through lateral transport. However,

more experimental evidence would be needed to support these conclusions based on structural analysis.

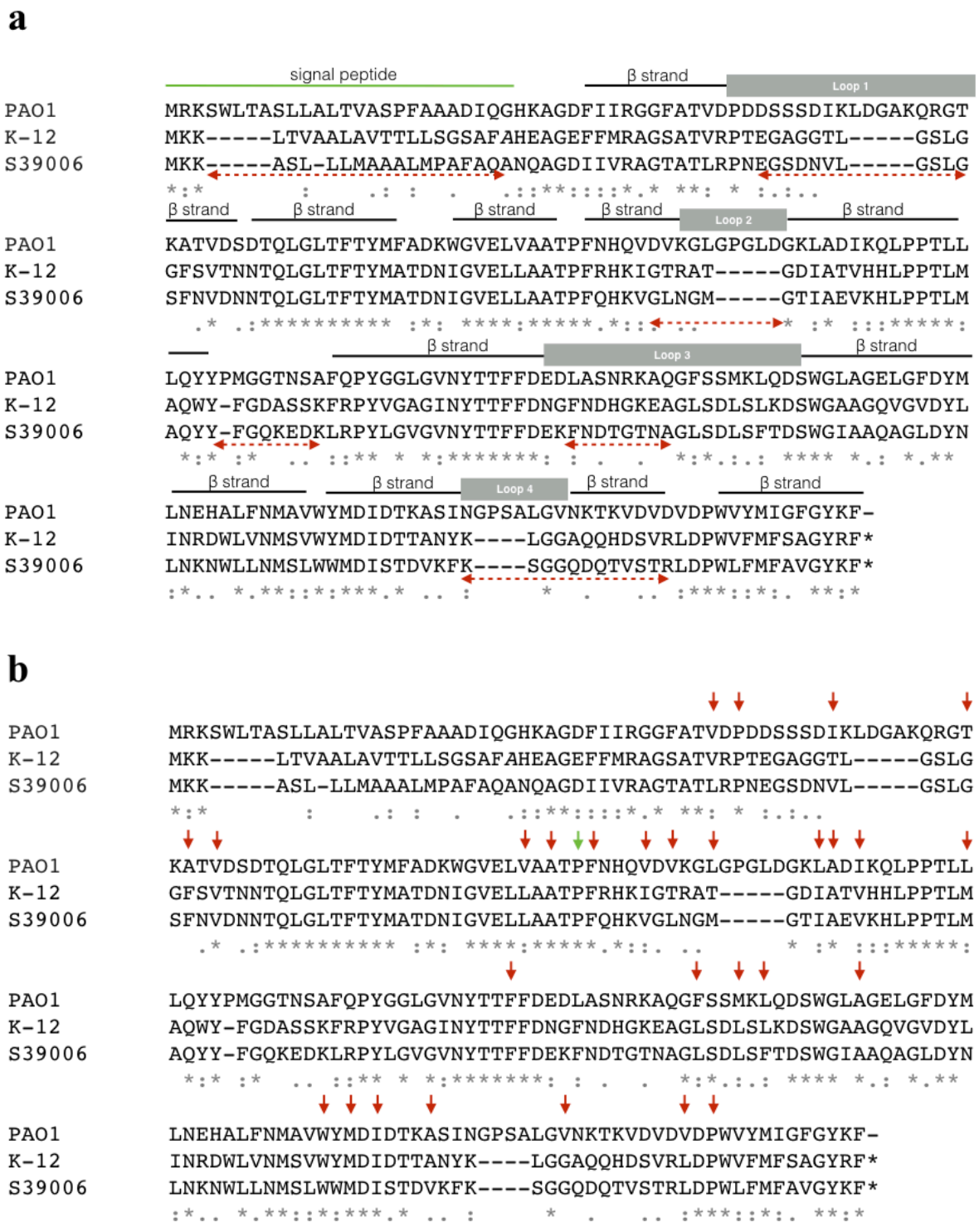


Figure 6.9: Protein sequence alignment of OmpW in *S*39006, *E.coli* K-12 and *P. aeruginosa* PAO1. Figure legend provided overleaf.

Figure 6.9: Protein sequence alignment of OmpW in *S*39006, *E.coli* K-12 and *P. aeruginosa* PAO1. "*" (asterisk): positions that have a single, fully conserved residue. ":" (colon): conservation between groups of strongly similar properties. "." (period): conservation between groups of weakly similar properties. (a) Prediction of the secondary structure of OmpW$_{S39006}$ based on the crystal structures of OmpW$_{K-12}$ and OmpW$_{PAO1}$. Short stretches of dissimilar sequences are mainly located in extracellular loop regions and labeled with dashed dotted bi-directional arrows in red color. (b) Conservation of functionally important residues in OmpW of the three different species. Red arrow: residues that are highly conserved among the three species and are located in the lumen of OmpW, facing inward. Green arrow: conserved Proline located at the lateral opening of OmpW in PAO1.

7. Discussion

The outer membrane (OM) of Gram-negative bacteria serves as a barrier between the cell and its external surroundings. OM consists of phospholipids, LPS and proteins, with 50% of the OM mass contributed by proteins (175, 176). Some of these outer membrane proteins (OMPs), also known as porins, form transmembrane β-barrels that control the selective permeability of certain molecules across the OM. Depending on the size of the proteins, the porins can be categorized into major porins (12-22 β-strands, trimeric) and minor porins (8 or 10 β-strands, monomeric) (177). Aside from performing their physiological function, some OMPs are exploited by phages as receptors for adsorption. For example, OmpC (Outer membrane protein C) is recognized by phage T4 (165, 178), LamB (Maltoporin) is recognized by phage λ (179), and BtuB (Vitamin B12 transporter) is recognized by some *S.* Typhimurium phages (22).

OmpW is a minor porin widely distributed in Gram-negative bacteria. In 2014, Xu and colleagues reported that OmpW acts as the receptor for *Vibrio cholerae* O1 El Tor phage VP5, making it the first instance where OmpW is associated with phage recognition (180). In this study, we have found that OmpW also acts as the receptor for phages ΦCBH8, ΦCHI14, ΦX20 and ΦCBH189, the first instance for *S*39006 phages. Even though highly conserved homologues of OmpW$_{S39006}$ exist in different species (Figure 6.3), ΦCBH8 was found to be highly specific for OmpW$_{S39006}$, failing to infect any other potential hosts tested. This is unsurprising because only one or a few key residues may be essential for phages recognition, therefore a slight difference in primary sequence could mean completely different interaction with the phage tail (181).

Despite the restricted host range of ΦCBH8, it was shown that expression of plasmid-encoded OmpW$_{S39006}$ protein enabled ΦCBH8 to infect a few otherwise non-susceptible hosts. Adsorption assays indicated, that upon induction of *ompW*$_{S39006}$ expression, ΦCBH8 was able to adsorb to these hosts rapidly, implying that receptor specificity, instead of other post-adsorption defense mechanisms, was the main obstacle for successful infection. In contrast, 13 other hosts that have also acquired the *ompW*$_{S39006}$-bearing plasmid failed to show susceptibility to ΦCBH8, the reason of which could be lack of *ompW*$_{S39006}$ expression from the plasmid, improper folding or transport of the exogenous OmpW$_{S39006}$ or the existence of other defense mechanisms. With the expanded host range of ΦCBH8, it was possible to test the phage for transduction across different species. ΦCBH8 transduced two plasmids with ColE1 origin in its original host and the new host, but failed

to transduce two chromosomal markers tested. The reason for this is unclear, but a possibility could be speculated from the requirement of headful packaging for generalized transduction. In order for a generalized transducing phage to "accidentally" package host DNA into its capsid, the size of the DNA fragment must be similar to the phage genome size (182, 183). Therefore, for ΦCBH8 to package *S*39006 DNA into its capsid, a chromosomal fragment bigger than 170 kb is required. Fragments of this size, equivalent to 3% of *S*39006 chromosome, might be non-existent by the time DNA packaging occurs since degradation of the host chromosome happens rapidly after initiation of infection. In contrast, studies have shown that plasmids are not necessarily degraded by phages during infection (184) and could therefore still be packaged into the phage capsid. The two plasmids tested in this study have a much smaller size than required for headful packaging, but studies have shown that ColE1-type plasmids such as pBR322 replicate via rolling circle replication and could form a concatemer large enough for headful packaging (185). Even though this explanation for ΦCBH8's ability to transduce plasmid and inability to transduce chromosomal markers remains to be experimentally verified, ΦCBH8 could still act as a useful biotechnological tool to transfer plasmids between hosts.

The discovery that OmpW acts as a phage receptor on the *S*39006 surface begged the question about the physiological function of OmpW. The answer to this question could help link *S*39006's susceptibility to phage with its other biological phenotypes. Literature search showed that no experimental data exists on the function of OmpW in *S*39006 or other *Serratia* species. So far, the only well-established function of OmpW is that of colicin S4 receptor in *E. coli* K-12. Other studies have proposed several functions for OmpW in different bacteria but are yet to be substantiated. For example, OmpW in *Vibrio cholerae* was proposed to transport carnitine (186), OmpW in *Salmonella typhimurium* was proposed to be an efflux pump of methyl viologen (187) and OmpW in *Acinetobacter baumannii* was found to transport iron (188). The transport function of OmpW is also predicted from structures of the protein in *E.coli* K-12 and *P. aeruginosa* PAO1 where the inner tunnel of OmpW is lined with hydrophobic residues and contains an opening in the periplasmic domain of the porin, suggesting that OmpW is involved in lateral transport of a small hydrophobic molecule. The proposed transport function of OmpW seems to serve the purpose of stress tolerance and adaptation. In the examples mentioned above, carnitine is a compatible solute important for osmo-adaption in the presence of high NaCl concentration, methyl viologen is the inducer of reactive oxygen species, and iron homeostasis is important for bacterial

survival. Furthermore, previous studies have also shown that OmpW is involved in bacterial resistance to antibiotics such as tetracycline and kanamycin (168). Aside from stress response and adaptation, OmpW is also reported to be important in virulence. OmpW in *E.coli* K-12 was found to protect bacteria against host phagocytosis in mice infection assays (170) while OmpW in *A. baumanni* acted as a highly immunogenic marker in mice infection (189).

The great variety of suggested functions for OmpW in different bacteria provided clues for its function in *S*39006, yet relevant assays carried out in this study comparing wild type *S*39006 and an OmpW mutant, looking at virulence, stress response and antibiotic resistance, did not yield any positive results. Furthermore, OmpW$_{S39006}$ was also found to be non-essential for *S*39006 when grown in LB and non-toxic when over-expressed in *E.coli* DH5α. Even though the number of phenotypes assayed in this study is far from exhaustive, the current negative results, most of which were obtained from growth in rich medium and optimal growth conditions, could imply that the function of OmpW might only be clear in certain conditions, such as when grown in certain nutrient sources, oxygen availability or temperature. In fact, bioinformatics prediction of potential regulators that may act on the *ompW*$_{S39006}$ promoter suggested that OmpW could be regulated by CRP, a global regulator involved in carbon metabolism, and FNR, a regulator involved in O_2 availability, consistent with data shown in *E.coli* K-12 (169). The *ompW* promoter also contains a SoxS binding site, consistent with data shown in *S. typhimurium* (190). Future work will focus on screening for regulators of *ompW*$_{S39006}$ to find clues on its function and investigate how OmpW regulation is linked to *S*39006 susceptibility to phage.

Finally, using the expanded host range of ΦCBH8, the host-dependency of ToxIN$_{Pa}$-mediated Abi activation was tested, which was the original aim in this study that led to the discovery of OmpW as a phage receptor. Using *E.coli* DH5α as a second host background where ToxIN$_{Pa}$ is known to be functional as an Abi system, it has been shown that ΦCBH8 is not aborted by ToxIN$_{Pa}$, in contrast to using *S*39006 as the host background. This host-dependency of Abi-activation by ΦCBH8 could imply that certain phage product(s) involved in ToxIN$_{Pa}$-sensitivity in *S*39006 is not expressed adequately or assembled properly in the new host background or growth conditions. It could also imply that Abi-activation is not directly through a phage product(s) but could involve certain responses from host-specific pathways acting as a "middleman". Nonetheless, the host-dependency observed in this study suggests that activation of ToxIN$_{Pa}$-mediated Abi is a very

specialized process that requires the correct combination of phage and host. This may be good news for phage therapy as a broad host range phage might not be “threatened” by Abi systems such as ToxIN$_{Pa}$ in one host even if it is aborted in another.

Chapter 7 – Final Discussion

ToxIN$_{Pa}$, the Type III TA system, was found on a cryptic plasmid in *P.atrosepticum* SCRI1039 and protected bacteria from phage infection through Abi, in its native host and in other bacteria (64). Despite understanding of its structure and mode of action on bacterial cells (64, 70, 90-92), little is known about the interaction between ToxIN$_{Pa}$ and phages, and how phage infection triggers the activation of abortive infection (Abi) mediated by ToxIN$_{Pa}$. This thesis describes work performed to characterize phages of *S*39006 that are sensitive to ToxIN$_{Pa}$ and their genetic loci that are involved in activating ToxIN$_{Pa}$-mediated Abi.

1. Summary of findings

Two phages isolated on *S*39006, ΦCBH8 and ΦCBH189, were shown to be ToxIN$_{Pa}$-sensitive. This is an addition to the ToxIN$_{Pa}$-sensitive *S*39006 phages isolated from the natural environment that includes ΦCHI14 and ΦX20 (105, 106). All four phages belonged to the *Myoviridae* family and were T4-like phages. Bioinformatic analysis showed that the four *S*39006 phages were highly similar to each other and formed a new genus, the closest relative of which was *E.coli* phage EcS1. Despite their high level of sequence similarity, ΦCBH8, ΦCHI14 and ΦX20 were able to produce spontaneous "escape" mutants of ToxIN$_{Pa}$ while ΦCBH189 could not. Escape mutants of ΦCBH8 were sequenced and their genomes compared to the wild type genome. ToxIN$_{Pa}$ escape loci of ΦCBH8 were characterized in detail and compared to those of ΦCHI14 and ΦX20.

ΦCBH8 could escape ToxIN$_{Pa}$ through two independent routes. Some ΦCBH8 mutants that survived in the presence of ToxIN$_{Pa}$ lost a large region in their genomes (large deletion, or LD mutants). Some ΦCBH8 mutants acquired mutations in the *asiA* gene. ΦCHI14 mutants also had large deletion mutations or *asiA* mutations while all ΦX20 mutants were LD mutants. The repeated occurrences of these escape loci highlighted their importance in ToxIN$_{Pa}$-sensitivity and the fact that these *S*39006 phages may possess a variety of modes of interaction with ToxIN$_{Pa}$.

The LD region lost in different LD mutants had different sizes and borders but shared a common core. The large size of the LD regions and the absence of these regions in other related phages suggested that the corresponding genes must be involved in some novel phage biology that called

for further characterization. The deletions probably occurred through recombination, based on the presence of direct repeats adjacent to the deletion borders. ΦCBH8o, the mutant with the smallest LD region of 6.5 kb, lost 13 *orf*s and 11 tRNAs. Despite exhaustive bioinformatics analysis, no predictable function could be assigned to any of the *orf*s. These *orf*s did not appear to be related to general phage fitness because ΦCBH8o showed no difference in adsorption efficiency or burst size compared to the wild type. Complementation assays of individual *orf*s deleted in ΦCBH8o showed that no *orf* expression could activate ToxIN$_{Pa}$-mediated Abi on its own, suggesting the simultaneous involvement of at least 2 non-contiguous *orf*s. The *orf*s might be located some distance apart from each other, presumably at the borders of the deleted regions, which would explain the large size of the deleted regions. Among the *orf*s deleted in ΦCBH8o, *orf*10 and *orf*18 were toxic to *S*39006 when expressed. *orf*10 induced filamentation of cells and *orf*18 could not be cloned without acquiring mutations. *orf*10 and *orf*18 were located at the borders of the common core deleted in all mutants, making them potential candidates responsible for Abi-activation. Furthermore, expression of *orf*16, *orf*17 and *orf*19 were also observed to partially reduce the titer of ΦCBH8o when expressed in the presence of ToxIN$_{Pa}$, making them likely candidates involved in Abi-activation.

Complementation with the wild type AsiA protein caused *asiA* mutant phages, previously insensitive to ToxIN$_{Pa}$, to be aborted by ToxIN$_{Pa}$ again. This suggested that AsiA plays an important role in activating ToxIN$_{Pa}$-mediated Abi. Most of the mutations occurred at the CTD of AsiA with the exception of a V13G mutant. Mutated AsiA protein was shown to have lost its function in inhibiting cell growth. Phages with mutated AsiA demonstrated decreased burst size and extended latent period. These results suggested that mutation of the AsiA protein affected its function in σ-appropriation, but enabled phages to escape ToxIN$_{Pa}$, therefore implying that the mechanism through which AsiA activates Abi may be related to σ-appropriation. Experimental evidence supported this view because AsiA was found to decrease expression from the *toxIN$_{Pa}$* promoter.

OmpW was identified as the receptor for the ToxIN$_{Pa}$-sensitive *S*39006 phages. Heterologous expression of *S*39006 OmpW in other hosts (*E.coli* DH5α, *D.solani* MK10 and *D.zeae* NCPP 3532) expanded the host range of ΦCBH8 and enabled assessment of ΦCBH8 sensitivity to ToxIN$_{Pa}$ in DH5α. Even though ΦCBH8 was strongly aborted by ToxIN$_{Pa}$ in *S*39006, it was not aborted in

DH5α where ToxIN$_{Pa}$ (expressed from the same vector) was known to be functional as an Abi system (64). This suggested that host background plays a role in activation of ToxIN$_{Pa}$-mediated Abi, at least in some strains. In addition to its sensitivity to ToxIN$_{Pa}$, ΦCBH8 was also found to have the capacity for transduction as it enabled the transfer of two different ColE1-type plasmids in *S*39006. Expansion of ΦCBH8 host range also enabled it to act as a transducing phage in DH5α. Finally, bioinformatic analysis suggested that OmpW could be the outer membrane porin that acts as a transporter of a small hydrophobic molecule related to metabolism or stress response.

2. Insights on the activation of ToxIN$_{Pa}$-mediated Abi

Given the results summarized above, two different models could be proposed for the activation of ToxIN$_{Pa}$-mediated Abi by ΦCBH8 - the "AsiA model" and the "toxicity model".

In the AsiA model (Figure 7.1a), the in-coming phage produces AsiA proteins to take over host transcription machineries. Through σ-appropriation, AsiA reduces transcription of host genes, including the *toxIN$_{Pa}$* genes carried on a plasmid. Reduced *toxIN$_{Pa}$* transcription would lead to inbalance of ToxI and ToxN and degradation of the more labile ToxI, causing the dissociation of the ToxIN$_{Pa}$ complex, and thereby "activation" of Abi. In addition to experimental evidence described in this study, the AsiA model is supported by biochemical data from previous studies where the ToxIN$_{Pa}$ complex was shown to be highly dynamic – ToxN degraded the full length ToxI transcript constantly *in vitro*, even in the presence of excess single ToxI repeats (91, 92). This highlighted the importance of ToxI:ToxN stoichiometry for ToxIN$_{Pa}$ stability and showed the consequence of transcription disruption by AsiA at the *toxIN$_{Pa}$* promoter, observed in this study. A similar mode of action was also proposed for phage T4 activating the phage exclusion function of the *hok/sok* system (88).

To further substantiate the AsiA model, future work could focus on comparing the cellular level of *toxIN$_{Pa}$* mRNA and comparing the ToxI:ToxN ratio in the presence or absence of AsiA. This could provide more solid proof that transcription disruption is the reason for Abi activation.

In the toxicity model (Figure 7.1b), the in-coming phage produces at least two proteins that are toxic to the host. These proteins trigger the activation of certain host response pathways that, in turn, activates ToxIN$_{Pa}$-mediated Abi through disruption of *toxIN$_{Pa}$* transcription, translation, or

direct interaction with the ToxIN$_{Pa}$ complex. Compared to the AsiA model, this model is less robust because while the LD region deleted in ΦCBH8 mutants was involved in Abi-activation, it

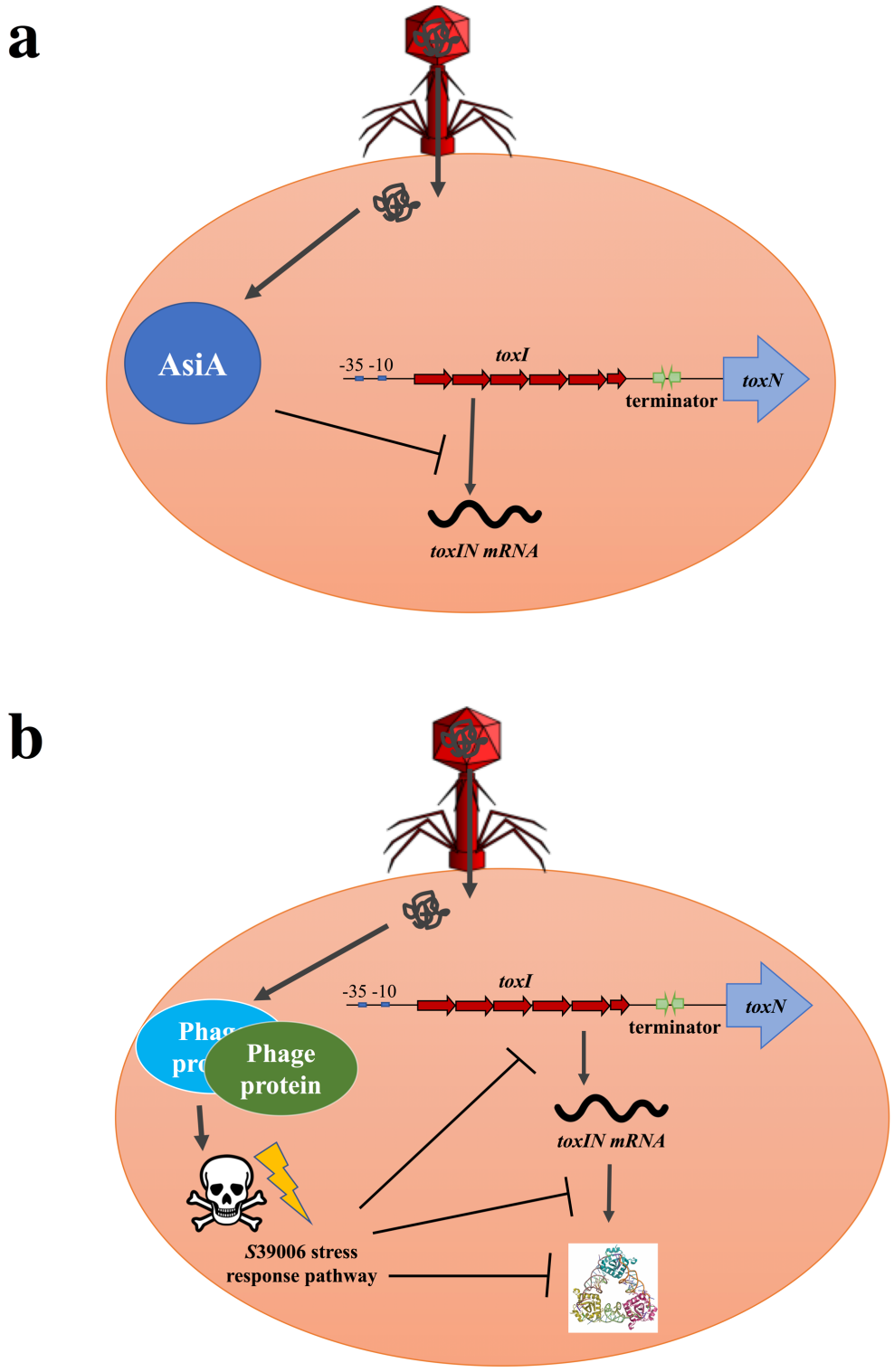


Figure 7.1: Two models proposed for how ΦCBH8 activates ToxIN$_{Pa}$-mediated Abi. (a) The AsiA model. (b) The toxicity model. Figure legend is provided overleaf.

Figure 7.1: Two models proposed for how ΦCBH8 activates ToxIN$_{Pa}$-mediated Abi. (a) The AsiA model. (b) The toxicity model. In the AsiA model, the in-coming phage produces AsiA proteins to take over host transcription machineries. Through σ-appropriation, AsiA reduces transcription of host genes, including the *toxIN$_{Pa}$* genes carried on a plasmid. Reduced *toxIN$_{Pa}$* transcription would lead to inbalance of ToxI and ToxN and degradation of the more labile ToxI, causing the dissociation of the ToxIN$_{Pa}$ complex, and thereby "activation" of Abi. In the toxicity model, the in-coming phage produces at least two proteins that are toxic to the host. These proteins trigger the activation of certain host response pathways that, in turn, activates ToxIN$_{Pa}$-mediated Abi through disruption of *toxIN$_{Pa}$* transcription, translation, or direct interaction with the ToxIN$_{Pa}$ complex.

was not possible in this study to narrow down the target to specific *orf*s. The presumed involvement of the toxic hypothetical proteins ORF10 and ORF18 in Abi-activation was only theoretical. However, this model was still suggested because of two reasons. First, protein M1-23 from ΦM1 infecting *P. atrosepticum*, a highly toxic protein, was involved in activating ToxIN$_{Pa}$-mediated Abi (98). Therefore, an analogy might be drawn between M1-23 and ORF10/ORF18. Second, ΦCBH8 was not aborted by ToxIN$_{Pa}$ in DH5α, highlighting the importance of host background, therefore suggesting that particular host pathways are involved.

Clearly, more extensive work needs to be done to support the toxicity model. First, to pinpoint specific *orf*s within the LD region that are involved in Abi-activation, *orf*s in the phage genome could be knocked out sequentially until ΦCBH8 becomes insensitive to ToxIN$_{Pa}$. Such targeted phage genome modification could be carried out with emerging techniques such CRISPR-Cas (191). Once target *orf*s are identified, the level and order of expression of these *orf*s could be tracked through transcriptomics, proteomics and metabolomics, through which additional phage *orf*(s) involved in the Abi-activation process could also be identified. This could be carried out in concert with monitoring the transcription, translation and stability of the ToxIN$_{Pa}$ complex to gather evidence on the interaction of phage *orf*(s) with ToxIN$_{Pa}$. In the meantime, the transcriptome, proteome and metabolome of the host could also be analyzed to identify any potentially specific host pathways involved in ToxIN$_{Pa}$-mediated Abi.

In summary, through analysis of *asiA* mutants and LD mutants of ΦCBH8, two distinct models of ToxIN$_{Pa}$-mediated Abi activation could be suggested. Of course, these two models might overlap since AsiA is also toxic to the cell. This reflects the flexibility of the ToxIN$_{Pa}$ system and its interaction with phage. The apparent flexibility and adaptability in ToxIN$_{Pa}$ system performance could also imply that ToxIN$_{Pa}$ has another unknown primary function(s) that responds to a wide variety of stimuli, and this may be one reason for the lack of an obvious, unified mechanism for the activation of ToxIN$_{Pa}$. It may be important, however, to remember that the original ToxIN$_{Pa}$ system was discovered in a culture collection strain of *Pectobacterium atrosepticum* where it was carried on a natural, but cryptic, plasmid replicon. So, by definition, all subsequent analyses of ToxIN$_{Pa}$ system functionality have been based on artificially engineered derivatives in different bacterial hosts, including *S*39006. Still, the flexibility of the ToxIN$_{Pa}$ system discovered in this

study and in previous studies (regardless of host background) echoes with previous hypothesis put forward by Wood and colleagues that aiding cellular response to a myriad of stresses is the primary function of TA systems (86). Indeed, ToxIN$_{Pa}$ system could be a good example of this hypothesis – it is readily activated by wide range of stimuli and signals. The activation by phages could be a by-product of this response system as some phage *orf*s could mimic the effect of signals related to stress and survival.

3. Other future work

Aside from the extensive work needed on phage escape loci and ToxIN$_{Pa}$ described above, research could also be carried out on other interesting aspects of phage biology uncovered as a result of this study. For example, aside from its role in Abi-activation, the *orf*s in the LD region of ΦCBH8 could be investigated in depth to explain their biological function, the reason for their existence in the phage genome and their evolutionary origin. If cellular elongation caused by ORF10 is due to its real function rather than an artifact caused by overexpression, then ORF10 could be exploited biotechnologically because of its potential ability to prevent septation in the cell.

Furthermore, the discovery that OmpW is exploited by various phages as a receptor to enter the host begs investigation of its natural physiological function. If the function and regulation of OmpW is understood, it would be possible to control *S*39006 susceptibility to phages by up- or down-regulating OmpW expression. Theoretically, this information may even prove useful in phage therapy studies and for biotechnological exploitation. Future work could focus on discovering the natural function of the OmpW porin through two different approaches. On one hand, more phenotypical assays could be carried out, especially testing equivalent phenotypes related to the presumed roles of OmpW in other bacteria. On the other hand, screening of *ompW* regulators could be attempted to gain more insight on the biological processes in which OmpW participates.

References

1. Twort FW, Lond LRCP. An investigation on the nature of ultra-microscopic viruses. The Lancet. 1915;Jan. 10th, 1914:101.
2. Duckworth DH. "Who discovered bacteriophage?". Bacteriol Rev. 1976;40(4):793-802.
3. Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH, Prangishvili D. Exploring the prokaryotic virosphere. Res Microbiol. 2008;159(5):306-13.
4. Keen EC. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays. 2015;37(1):6-9.
5. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, et al. Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*. 2017(15):161 – 8.
6. Ackermann H-W. Phage Classification and Characterization. In: Clokie MRJ, Kropinski AM, editors. Bacteriophages: Methods and Protocols. Volume 1: Isolation, Characterization, and Interactions Humana Press; 2009. p. 127-54.
7. Bradley DE. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev. 1967;31(4):230-314.
8. Maniloff J, Ackermann HW. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order Caudovirales. Arch Virol. 1998;143(10):2051-63.
9. Ackermann HW. Phage classification and characterization. Methods Mol Biol. 2009;501:127-40.
10. Yap ML, Rossmann MG. Structure and function of bacteriophage T4. Future Microbiol. 2014;9(12):1319-27.
11. Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG. Structure and morphogenesis of bacteriophage T4. Cell Mol Life Sci. 2003;60(11):2356-70.
12. Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J, Vogensen FK, Sullivan MB, et al. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol. 2016;161(4):1095-9.
13. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun. 2014;5:4498.
14. Lima-Mendez G, Toussaint A, Leplae R. Analysis of the phage sequence space: the benefit of structured information. Virology. 2007;365(2):241-9.
15. Cenens W, Makumi A, Mebrhatu MT, Lavigne R, Aertsen A. Phage-host interactions during pseudolysogeny: Lessons from the Pid/dgo interaction. Bacteriophage. 2013;3(1):e25029.
16. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363(4).
17. Gaidelyte A, Cvirkaite-Krupovic V, Daugelavicius R, Bamford JK, Bamford DH. The entry mechanism of membrane-containing phage Bam35 infecting Bacillus thuringiensis. J Bacteriol. 2006;188(16):5925-34.

18. Munsch-Alatossava P, Alatossava T. The extracellular phage-host interactions involved in the bacteriophage LL-H infection of Lactobacillus delbrueckii ssp. lactis ATCC 15808. Front Microbiol. 2013;4:408.
19. Davison S, Couture-Tosi E, Candela T, Mock M, Fouet A. Identification of the Bacillus anthracis (gamma) phage receptor. J Bacteriol. 2005;187(19):6742-9.
20. Meadow PM, Wells PL. Receptor Sites for R-type Pyocins and Bacteriophage E79 in the Core Part of the Lipopolysaccharide of Pseudomonas aeruginosa PAC 1. Microbiology. 1978;108:339-43.
21. Ho TD, Slauch JM. OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of Salmonella. J Bacteriol. 2001;183(4):1495-8.
22. Shin H, Lee JH, Kim H, Choi Y, Heu S, Ryu S. Receptor diversity and host interaction of bacteriophages infecting Salmonella enterica serovar Typhimurium. PLoS One. 2012;7(8):e43392.
23. Bae HW, Cho YH. Complete Genome Sequence of Pseudomonas aeruginosa Podophage MPK7, Which Requires Type IV Pili for Infection. Genome Announc. 2013;1(5).
24. Pickard D, Toribio AL, Petty NK, van Tonder A, Yu L, Goulding D, et al. A conserved acetyl esterase domain targets diverse bacteriophages to the Vi capsular receptor of Salmonella enterica serovar Typhi. J Bacteriol. 2010;192(21):5746-54.
25. Hinton DM. Transcriptional control in the prereplicative phase of T4 development. Virol J. 2010;7:289.
26. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev. 2003;67(1):86-156, table of contents.
27. Keen EC. Paradigms of pathogenesis: targeting the mobile genetic elements of disease. Front Cell Infect Microbiol. 2012;2:161.
28. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. Prophage genomics. Microbiol Mol Biol Rev. 2003;67(2):238-76, table of contents.
29. R C. The Bacteriophages: Oxford University Press; 2006.
30. Young R. Phage lysis: three steps, three choices, one outcome. J Microbiol. 2014;52(3):243-58.
31. HERSHEY AD, CHASE M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. 1952;36(1):39-56.
32. Salmond GP, Fineran PC. A century of the phage: past, present and future. Nat Rev Microbiol. 2015;13(12):777-86.
33. Monson R, Smith DS, Matilla MA, Roberts K, Richardson E, Drew A, et al. A Plasmid-Transposon Hybrid Mutagenesis System Effective in a Broad Range of Enterobacteria. Front Microbiol. 2015;6:1442.
34. Nathans D, Smith HO. Restriction endonucleases in the analysis and restructuring of dna molecules. Annu Rev Biochem. 1975;44:273-93.
35. Zhu B. Bacteriophage T7 DNA polymerase - sequenase. Front Microbiol. 2014;5:181.
36. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science. 1985;228(4705):1315-7.
37. Haq IU, Chaudhry WN, Akhtar MN, Andleeb S, Qadri I. Bacteriophages and their implications on future biotechnology: a review. Virol J. 2012;9:9.

38. Campbell A. The future of bacteriophage biology. Nat Rev Genet. 2003;4(6):471-7.
39. Clark JR, March JB. Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. Trends Biotechnol. 2006;24(5):212-8.
40. Dublanchet A, Bourne S. The epic of phage therapy. Can J Infect Dis Med Microbiol. 2007;18(1):15-8.
41. Frenkel D, Solomon B. Filamentous phage as vector-mediated antibody delivery to the brain. Proc Natl Acad Sci U S A. 2002;99(8):5675-9.
42. Seed KD. Battling Phages: How Bacteria Defend against Viral Attack. PLoS Pathog. 2015;11(6):e1004847.
43. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science. 2002;295(5562):2091-4.
44. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol. 2010;8(5):317-27.
45. Riede I, Eschbach ML. Evidence that TraT interacts with OmpA of Escherichia coli. FEBS Lett. 1986;205(2):241-5.
46. Stummeyer K, Schwarzer D, Claus H, Vogel U, Gerardy-Schahn R, Mühlenhoff M. Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages. Mol Microbiol. 2006;60(5):1123-35.
47. Sutherland IW. Polysaccharide lyases. FEMS Microbiol Rev. 1995;16(4):323-47.
48. Bebeacua C, Lorenzo Fajardo JC, Blangy S, Spinelli S, Bollmann S, Neve H, et al. X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target. Mol Microbiol. 2013;89(1):152-65.
49. Enikeeva FN, Severinov KV, Gelfand MS. Restriction-modification systems and bacteriophage invasion: who wins? J Theor Biol. 2010;266(4):550-9.
50. Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature. 2015;526(7571):55-61.
51. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327(5962):167-70.
52. Chaudhary K. BacteRiophage EXclusion (BREX): A novel anti-phage mechanism in the arsenal of bacterial defense system. J Cell Physiol. 2018;233(2):771-3.
53. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, et al. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. 2015;34(2):169-83.
54. Barrangou R, van der Oost J. Bacteriophage exclusion, a new defense system. EMBO J. 2015;34(2):134-5.
55. Ram G, Chen J, Ross HF, Novick RP. An insight into staphylococcal pathogenicity island-mediated interference with phage late gene transcription. Bacteriophage. 2015;5(2):e1028608.
56. Novick RP, Subedi A. The SaPIs: mobile pathogenicity islands of Staphylococcus. Chem Immunol Allergy. 2007;93:42-57.
57. Forde A, Fitzgerald GF. Bacteriophage defence systems in lactic acid bacteria. Antonie Van Leeuwenhoek. 1999;76(1-4):89-113.

58. Chopin MC, Chopin A, Bidnenko E. Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol. 2005;8(4):473-9.
59. Curtis FA, Reed P, Sharples GJ. Evolution of a phage RuvC endonuclease for resolution of both Holliday and branched DNA junctions. Mol Microbiol. 2005;55(5):1332-45.
60. Bouchard JD, Moineau S. Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. J Bacteriol. 2004;186(11):3649-52.
61. Durmaz E, Klaenhammer TR. Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. J Bacteriol. 2007;189(4):1417-25.
62. Dy RL, Przybilski R, Semeijn K, Salmond GP, Fineran PC. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res. 2014;42(7):4590-605.
63. Samson JE, Spinelli S, Cambillau C, Moineau S. Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. Mol Microbiol. 2013;87(4):756-68.
64. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A. 2009;106(3):894-9.
65. Yamaguchi Y, Park JH, Inouye M. Toxin-antitoxin systems in bacteria and archaea. Annu Rev Genet. 2011;45:61-79.
66. Hall AM, Gollan B, Helaine S. Toxin-antitoxin systems: reversible toxicity. Curr Opin Microbiol. 2017;36:102-10.
67. Shidore T, Triplett LR. Toxin-Antitoxin Systems: Implications for Plant Disease. Annu Rev Phytopathol. 2017;55:161-79.
68. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat Chem Biol. 2016;12(4):208-14.
69. Unterholzner SJ, Poppenberger B, Rozhon W. Toxin-antitoxin systems: Biology, identification, and application. Mob Genet Elements. 2013;3(5):e26219.
70. Blower TR, Pei XY, Short FL, Fineran PC, Humphreys DP, Luisi BF, et al. A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat Struct Mol Biol. 2011;18(2):185-90.
71. Ogura T, Hiraga S. Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc Natl Acad Sci U S A. 1983;80(15):4784-8.
72. Gerdes K, Rasmussen PB, Molin S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proc Natl Acad Sci U S A. 1986;83(10):3116-20.
73. Yarmolinsky MB. Programmed cell death in bacterial populations. Science. 1995;267(5199):836-7.
74. Van Melderen L, Saavedra De Bast M. Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet. 2009;5(3):e1000437.
75. Song S, Wood TK. Post-segregational Killing and Phage Inhibition Are Not Mediated by Cell Death Through Toxin/Antitoxin Systems. Front Microbiol. 2018;9:814.

76. Moyed HS, Bertrand KP. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. J Bacteriol. 1983;155(2):768-75.
77. Masuda H, Inouye M. Toxins of Prokaryotic Toxin-Antitoxin Systems with Sequence-Specific Endoribonuclease Activity. Toxins (Basel). 2017;9(4).
78. Maisonneuve E, Castro-Camargo M, Gerdes K. (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. Cell. 2013;154(5):1140-50.
79. Maisonneuve E, Castro-Camargo M, Gerdes K. (p)ppGpp Controls Bacterial Persistence by Stochastic Induction of Toxin-Antitoxin Activity. Cell. 2018;172(5):1135.
80. Stewart PS, Franklin MJ. Physiological heterogeneity in biofilms. Nat Rev Microbiol. 2008;6(3):199-210.
81. Lewis K. Persister cells. Annu Rev Microbiol. 2010;64:357-72.
82. Wen Y, Behiels E, Devreese B. Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. Pathog Dis. 2014;70(3):240-9.
83. Wang X, Kim Y, Hong SH, Ma Q, Brown BL, Pu M, et al. Antitoxin MqsA helps mediate the bacterial general stress response. Nat Chem Biol. 2011;7(6):359-66.
84. Christensen SK, Mikkelsen M, Pedersen K, Gerdes K. RelE, a global inhibitor of translation, is activated during nutritional stress. Proc Natl Acad Sci U S A. 2001;98(25):14328-33.
85. Kwan BW, Lord DM, Peti W, Page R, Benedik MJ, Wood TK. The MqsR/MqsA toxin/antitoxin system protects Escherichia coli during bile acid stress. Environ Microbiol. 2015;17(9):3168-81.
86. Soo VW, Cheng HY, Kwan BW, Wood TK. de novo synthesis of a bacterial toxin/antitoxin system. Sci Rep. 2014;4:4807.
87. Blower TR, Short FL, Rao F, Mizuguchi K, Pei XY, Fineran PC, et al. Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. Nucleic Acids Res. 2012;40(13):6158-73.
88. Pecota DC, Wood TK. Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J Bacteriol. 1996;178(7):2044-50.
89. Hazan R, Engelberg-Kulka H. Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Genet Genomics. 2004;272(2):227-34.
90. Blower TR, Fineran PC, Johnson MJ, Toth IK, Humphreys DP, Salmond GP. Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of Erwinia. J Bacteriol. 2009;191(19):6029-39.
91. Short FL, Pei XY, Blower TR, Ong SL, Fineran PC, Luisi BF, et al. Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. Proc Natl Acad Sci U S A. 2013;110(3):E241-9.
92. Short FL, Akusobi C, Broadhurst WR, Salmond GPC. The bacterial Type III toxin-antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent antiviral abortive infection activity. Sci Rep. 2018;8(1):1013.
93. Rao F, Short FL, Voss JE, Blower TR, Orme AL, Whittaker TE, et al. Co-evolution of quaternary organization and novel RNA tertiary interactions revealed in the crystal

structure of a bacterial protein-RNA toxin-antitoxin system. Nucleic Acids Res. 2015;43(19):9529-40.
94. Bélanger M, Moineau S. Mutational Analysis of the Antitoxin in the Lactococcal Type III Toxin-Antitoxin System AbiQ. Appl Environ Microbiol. 2015;81(11):3848-55.
95. Rao F. Structural and Functional Characterisation of novel Type III systems: University of Cambridge; 2014.
96. Short FL, Monson RE, Salmond GP. A Type III protein-RNA toxin-antitoxin system from Bacillus thuringiensis promotes plasmid retention during spore development. RNA Biol. 2015;12(9):933-7.
97. Emond E, Dion E, Walker SA, Vedamuthu ER, Kondo JK, Moineau S. AbiQ, an abortive infection mechanism from Lactococcus lactis. Appl Environ Microbiol. 1998;64(12):4748-56.
98. Blower TR, Chai R, Przybilski R, Chindhy S, Fang X, Kidman SE, et al. Evolution of Pectobacterium Bacteriophage ΦM1 To Escape Two Bifunctional Type III Toxin-Antitoxin and Abortive Infection Systems through Mutations in a Single Viral Gene. Appl Environ Microbiol. 2017;83(8).
99. Samson JE, Bélanger M, Moineau S. Effect of the abortive infection mechanism and type III toxin/antitoxin system AbiQ on the lytic cycle of Lactococcus lactis phages. J Bacteriol. 2013;195(17):3947-56.
100. Blower TR, Evans TJ, Przybilski R, Fineran PC, Salmond GP. Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. PLoS Genet. 2012;8(10):e1003023.
101. Parker WL, Rathnum ML, Wells JS, Trejo WH, Principe PA, Sykes RB. SQ 27,860, a simple carbapenem produced by species of Serratia and Erwinia. J Antibiot (Tokyo). 1982;35(6):653-60.
102. Coulthurst SJ, Barnard AM, Salmond GP. Regulation and biosynthesis of carbapenem antibiotics in bacteria. Nat Rev Microbiol. 2005;3(4):295-306.
103. Williamson NR, Fineran PC, Leeper FJ, Salmond GP. The biosynthesis and regulation of bacterial prodiginines. Nat Rev Microbiol. 2006;4(12):887-99.
104. Fineran PC, Iglesias Cans MC, Ramsay JP, Wilf NM, Cossyleon D, McNeil MB, et al. Draft Genome Sequence of Serratia sp. Strain ATCC 39006, a Model Bacterium for Analysis of the Biosynthesis and Regulation of Prodigiosin, a Carbapenem, and Gas Vesicles. Genome Announc. 2013;1(6).
105. Akusobi C. Characterization of Bacteriophage ΦCHI14 and its $ToxIN_{Pa}$ resistance loci. Cambridge: University of Cambridge; 2013.
106. Fang X. Characterization of new enterobacterial phages and their responses to the $ToxIN_{Pa}$ system. Cambridge: University of Cambridge; 2014.
107. Herrero M, de Lorenzo V, Timmis KN. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. J Bacteriol. 1990;172(11):6557-67.
108. Bainton NJ, Stead P, Chhabra SR, Bycroft BW, Salmond GP, Stewart GS, et al. N-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in Erwinia carotovora. Biochem J. 1992;288 (Pt 3):997-1004.
109. Demarre G, Guérout AM, Matsumoto-Mashimo C, Rowe-Magnus DA, Marlière P, Mazel D. A new family of mobilizable suicide plasmids based on broad host range R388

plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate Escherichia coli host strains. Res Microbiol. 2005;156(2):245-55.
110. Thomson NR, Cox A, Bycroft BW, Stewart GS, Williams P, Salmond GP. The rap and hor proteins of Erwinia, Serratia and Yersinia: a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in bacterial pathogens. Mol Microbiol. 1997;26(3):531-44.
111. Coulthurst SJ, Williamson NR, Harris AK, Spring DR, Salmond GP. Metabolic and regulatory engineering of Serratia marcescens: mimicking phage-mediated horizontal acquisition of antibiotic biosynthesis and quorum-sensing capacities. Microbiology. 2006;152(Pt 7):1899-911.
112. Matilla MA, Drew A, Udaondo Z, Krell T, Salmond GP. Genome Sequence of Serratia plymuthica A153, a Model Rhizobacterium for the Investigation of the Synthesis and Regulation of Haterumalides, Zeamine, and Andrimid. Genome Announc. 2016;4(3).
113. Pritchard L, Humphris S, Baeyen S, Maes M, Van Vaerenbergh J, Elphinstone J, et al. Draft Genome Sequences of Four Dickeya dianthicola and Four Dickeya solani Strains. Genome Announc. 2013;1(4).
114. Grenier AM, Duport G, Pagès S, Condemine G, Rahbé Y. The phytopathogen Dickeya dadantii (Erwinia chrysanthemi 3937) is a pathogen of the pea aphid. Appl Environ Microbiol. 2006;72(3):1956-65.
115. Pritchard L, Humphris S, Saddler GS, Elphinstone JG, Pirhonen M, Toth IK. Draft genome sequences of 17 isolates of the plant pathogenic bacterium dickeya. Genome Announc. 2013;1(6).
116. Williamson NR, Simonsen HT, Ahmed RA, Goldet G, Slater H, Woodley L, et al. Biosynthesis of the red antibiotic, prodigiosin, in Serratia: identification of a novel 2-methyl-3-n-amyl-pyrrole (MAP) assembly pathway, definition of the terminal condensing enzyme, and implications for undecylprodigiosin biosynthesis in Streptomyces. Mol Microbiol. 2005;56(4):971-89.
117. Bell KS, Sebaihia M, Pritchard L, Holden MT, Hyman LJ, Holeva MC, et al. Genome sequence of the enterobacterial phytopathogen Erwinia carotovora subsp. atroseptica and characterization of virulence factors. Proc Natl Acad Sci U S A. 2004;101(30):11105-10.
118. Flyg C, Kenne K, Boman HG. Insect pathogenic properties of Serratia marcescens: phage-resistant mutants with a decreased resistance to Cecropia immunity and a decreased virulence to Drosophila. J Gen Microbiol. 1980;120(1):173-81.
119. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature. 2000;406(6799):959-64.
120. Evans TJ, Crow MA, Williamson NR, Orme W, Thomson NR, Komitopoulou E, et al. Characterization of a broad-host-range flagellum-dependent phage that mediates high-efficiency generalized transduction in, and between, Serratia and Pantoea. Microbiology. 2010;156(Pt 1):240-7.
121. Petty NK, Toribio AL, Goulding D, Foulds I, Thomson N, Dougan G, et al. A generalized transducing phage for the murine pathogen Citrobacter rodentium. Microbiology. 2007;153(Pt 9):2984-8.

122. Sambrook. Molecular cloning; a laboratory manual. 2nd Edition ed. *New York*, *New York*, *USA*: *Cold Spring Harbour Laboratory Press*; 1989.
123. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995;177(14):4121-30.
124. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, et al. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene. 1977;2(2):95-113.
125. Blower TR. Functional and structural studies of the toxIN abortive infection and toxin-antitoxin locus from *Erwinia carotovora* subspecies *atroseptica* University of Cambridge; 2009.
126. Short FL. Assembly, specificity and function of prokaryotic Type III protein-RNA toxin-antitoxin systems: University of Cambridge; 2013.
127. Denes T, den Bakker HC, Tokman JI, Guldimann C, Wiedmann M. Selection and Characterization of Phage-Resistant Mutant Strains of Listeria monocytogenes Reveal Host Genes Linked to Phage Adsorption. Appl Environ Microbiol. 2015;81(13):4295-305.
128. Lee CM, Monson RE, Adams RM, Salmond GPC. The LacI-Family Transcription Factor, RbsR, Is a Pleiotropic Regulator of Motility, Virulence, Siderophore and Antibiotic Production, Gas Vesicle Morphogenesis and Flotation in. Front Microbiol. 2017;8:1678.
129. Sousa SA, Ramos CG, Moreira LM, Leitão JH. The hfq gene is required for stress resistance and full virulence of Burkholderia cepacia to the nematode Caenorhabditis elegans. Microbiology. 2010;156(Pt 3):896-908.
130. Lundberg BE, Wolf RE, Dinauer MC, Xu Y, Fang FC. Glucose 6-phosphate dehydrogenase is required for Salmonella typhimurium virulence and resistance to reactive oxygen and nitrogen intermediates. Infect Immun. 1999;67(1):436-8.
131. Fineran PC, Williamson NR, Lilley KS, Salmond GP. Virulence and prodigiosin antibiotic biosynthesis in Serratia are regulated pleiotropically by the GGDEF/EAL domain protein, PigX. J Bacteriol. 2007;189(21):7653-62.
132. Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 1998;26(4):1107-15.
133. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999;27(23):4636-41.
134. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725-38.
135. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955-64.
136. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11-6.
137. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 2010;10:8.
138. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36(Web Server issue):W465-9.
139. Rao F. Structural and functional characterisation of novel Type III systems. Cambridge: University of Cambridge; 2014.

140. Blower TR, Short FL, Fineran PC, Salmond GP. Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication. Bacteriophage. 2012;2(4):234-8.
141. Adriaenssens E, Brister JR. How to Name and Classify Your Phage: An Informal Guide. Viruses. 2017;9(4).
142. Comeau AM, Krisch HM. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol Biol Evol. 2008;25(7):1321-32.
143. Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD. Genetic diversity among five T4-like bacteriophages. Virol J. 2006;3:30.
144. Potapov S, Belykh O, Krasnopeev A, Gladkikh A, Kabilov M, Tupikin A, et al. Assessing the diversity of the g23 gene of T4-like bacteriophages from Lake Baikal with high-throughput sequencing. FEMS Microbiol Lett. 2018;365(3).
145. Adriaenssens EM, Cowan DA. Using signature genes as tools to assess environmental viral ecology and diversity. Appl Environ Microbiol. 2014;80(15):4470-80.
146. Singer BS, Westlye J. Deletion formation in bacteriophage T4. J Mol Biol. 1988;202(2):233-43.
147. Pierce JC, Masker W. Genetic deletions between directly repeated sequences in bacteriophage T7. Mol Gen Genet. 1989;217(2-3):215-22.
148. Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, et al. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Sci Rep. 2016;6:34338.
149. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.
150. Delattre H, Souiai O, Fagoonee K, Guerois R, Petit MA. Phagonaute: A web-based interface for phage synteny browsing and protein function prediction. Virology. 2016;496:42-50.
151. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages. Genome Res. 2007;17(10):1486-95.
152. Okada K. Physical map of the dispensable region of the genome of E. coli bacteriophage BF23. Gene. 1980;8(4):369-90.
153. Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J, Hatfull GF. Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. Mol Microbiol. 2013;88(3):577-89.
154. Petrov VM, Ratnayaka S, Nolan JM, Miller ES, Karam JD. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Virol J. 2010;7:292.
155. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev. 2004;17(1):14-56.
156. Karaolis DK, Somara S, Maneval DR, Johnson JA, Kaper JB. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature. 1999;399(6734):375-9.
157. Yuan AH, Nickels BE, Hochschild A. The bacteriophage T4 AsiA protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A. 2009;106(16):6597-602.

158. Hinton DM, Pande S, Wais N, Johnson XB, Vuthoori M, Makela A, et al. Transcriptional takeover by sigma appropriation: remodelling of the sigma70 subunit of Escherichia coli RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. Microbiology. 2005;151(Pt 6):1729-40.
159. Minakhin L, Severinov K. Transcription regulation by bacteriophage T4 AsiA. Protein Expr Purif. 2005;41(1):1-8.
160. Nyström T. Stationary-phase physiology. Annu Rev Microbiol. 2004;58:161-81.
161. Gallet R, Kannoly S, Wang IN. Effects of bacteriophage traits on plaque formation. BMC Microbiol. 2011;11:181.
162. Sharma UK, Praveen PV, Balganesh TS. Mutational analysis of bacteriophage T4 AsiA: involvement of N- and C-terminal regions in binding to sigma(70) of Escherichia coli in vivo. Gene. 2002;295(1):125-34.
163. Ouhammouch M, Orsini G, Brody EN. The asiA gene product of bacteriophage T4 is required for middle mode RNA synthesis. J Bacteriol. 1994;176(13):3956-65.
164. Toth IK, Mulholland V, Cooper V, Bentley S, Shih Y-L, Perombelon MCM, et al. Generalized transduction in the potato
blackleg pathogen *Erwinia carotovora* subsp.
atroseptica by bacteriophage M1. Microbiology. 1997;143:2433-8.
165. Chen M, Zhang L, Abdelgader SA, Yu L, Xu J, Yao H, et al. Alterations in gp37 expand the host range of a T4-like phage. Appl Environ Microbiol. 2017.
166. Mahichi F, Synnott AJ, Yamamichi K, Osada T, Tanji Y. Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. FEMS Microbiol Lett. 2009;295(2):211-7.
167. Pilsl H, Smajs D, Braun V. Characterization of colicin S4 and its receptor, OmpW, a minor protein of the Escherichia coli outer membrane. J Bacteriol. 1999;181(11):3578-81.
168. Xu C, Lin X, Ren H, Zhang Y, Wang S, Peng X. Analysis of outer membrane proteome of Escherichia coli related to resistance to ampicillin and tetracycline. Proteomics. 2006;6(2):462-73.
169. Xiao M, Lai Y, Sun J, Chen G, Yan A. Transcriptional Regulation of the Outer Membrane Porin Gene ompW Reveals its Physiological Role during the Transition from the Aerobic to the Anaerobic Lifestyle of Escherichia coli. Front Microbiol. 2016;7:799.
170. Wu XB, Tian LH, Zou HJ, Wang CY, Yu ZQ, Tang CH, et al. Outer membrane protein OmpW of Escherichia coli is required for resistance to phagocytosis. Res Microbiol. 2013;164(8):848-55.
171. Bott M. Anaerobic citrate metabolism and its regulation in enterobacteria. Arch Microbiol. 1997;167(2/3):78-88.
172. Hong H, Patel DR, Tamm LK, van den Berg B. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. J Biol Chem. 2006;281(11):7568-77.
173. Touw DS, Patel DR, van den Berg B. The crystal structure of OprG from Pseudomonas aeruginosa, a potential channel for transport of hydrophobic molecules across the outer membrane. PLoS One. 2010;5(11):e15016.

174. Horst R, Stanczak P, Wüthrich K. NMR polypeptide backbone conformation of the E. coli outer membrane protein W. Structure. 2014;22(8):1204-9.
175. van der Ley P, Heckels JE, Virji M, Hoogerhout P, Poolman JT. Topology of outer membrane porins in pathogenic Neisseria spp. Infect Immun. 1991;59(9):2963-71.
176. Koebnik R, Locher KP, Van Gelder P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol. 2000;37(2):239-53.
177. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev. 2003;67(4):593-656.
178. Yu F, Mizushima S. Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4. J Bacteriol. 1982;151(2):718-22.
179. Randall-Hazelbauer L, Schwartz M. Isolation of the bacteriophage lambda receptor from Escherichia coli. J Bacteriol. 1973;116(3):1436-46.
180. Xu D, Zhang J, Liu J, Xu J, Zhou H, Zhang L, et al. Outer membrane protein OmpW is the receptor for typing phage VP5 in the Vibrio cholerae O1 El Tor biotype. J Virol. 2014;88(12):7109-11.
181. Charbit A, Werts C, Michel V, Klebba PE, Quillardet P, Hofnung M. A role for residue 151 of LamB in bacteriophage lambda adsorption: possible steric effect of amino acid substitutions. J Bacteriol. 1994;176(11):3204-9.
182. Ozeki H, Ikeda H, 0.0px ppmppp, Helvetica} fp. Transduction Mechanisms. Annual Review of Genetics. 1968;2:245-78.
183. Ubelaker MH, Rosenblum ED. Transduction of plasmid determinants in Staphylococcus aureus and Escherichia coli. J Bacteriol. 1978;133(2):699-707.
184. Keen EC, Bliskovsky VV, Malagon F, Baker JD, Prince JS, Klaus JS, et al. Novel "Superspreader" Bacteriophages Promote Horizontal Gene Transfer by Transformation. MBio. 2017;8(1).
185. Mann BA, Slauch JM. Transduction of low-copy number plasmids by bacteriophage P22. Genetics. 1997;146(2):447-56.
186. Fu X, Zhang J, Li T, Zhang M, Li J, Kan B. The Outer Membrane Protein OmpW Enhanced. Front Microbiol. 2017;8:2703.
187. Gil F, Ipinza F, Fuentes J, Fumeron R, Villarreal JM, Aspée A, et al. The ompW (porin) gene mediates methyl viologen (paraquat) efflux in Salmonella enterica serovar typhimurium. Res Microbiol. 2007;158(6):529-36.
188. Catel-Ferreira M, Marti S, Guillon L, Jara L, Coadou G, Molle V, et al. The outer membrane porin OmpW of Acinetobacter baumannii is involved in iron uptake and colistin binding. FEBS Lett. 2016;590(2):224-31.
189. Huang W, Wang S, Yao Y, Xia Y, Yang X, Long Q, et al. OmpW is a potential target for eliciting protective immunity against Acinetobacter baumannii infections. Vaccine. 2015;33(36):4479-85.
190. Gil F, Hernández-Lucas I, Polanco R, Pacheco N, Collao B, Villarreal JM, et al. SoxS regulates the expression of the Salmonella enterica serovar Typhimurium ompW gene. Microbiology. 2009;155(Pt 8):2490-7.
191. Lemay ML, Tremblay DM, Moineau S. Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. ACS Synth Biol. 2017;6(7):1351-8.

Appendix I - Genome annotation of ΦCBH8

Strand	Gene number	Start position	Stop position	Gene length (b.p.)	Product length (a.a.)	Predicted gene product
reverse	1	2169	1	2169	722	rIIA - protector from prophage-induced early lysis
reverse	2	2383	2180	204	67	hypothetical protein
reverse	3	2705	2370	336	111	hypothetical protein
reverse	4	2924	2772	153	50	hypothetical protein
reverse	5	4806	2968	1839	612	DNA topoisomerase II large subunit
reverse	6	5120	4872	249	82	conserved hypothetical protein
reverse	7	5299	5120	180	59	fmdD - family putative regulatory protein
reverse	8	5624	5286	339	112	hypothetical protein
reverse	9	5964	5614	351	116	hypothetical protein
reverse	10	6404	6048	357	118	hypothetical protein
reverse	11	6631	6404	228	75	cef - modifier of suppressor tRNAs
reverse	12	7099	6701	399	132	hypothetical protein
reverse	13	7773	7117	657	218	phage-associated homing endonuclease
reverse	14	8453	7770	684	227	dexA exonuclease A
reverse	15	8679	8446	234	77	hypothetical protein
reverse	16	10012	8687	1326	441	RecD-like DNA helicase
reverse	17	10308	10012	297	98	conserved hypothetical protein
reverse	18	11003	10308	696	231	putative anti-sigma factor
reverse	19	11698	11090	609	202	ADP-ribosylase
reverse	20	11935	11753	183	60	hypothetical protein
reverse	21	12060	11935	126	41	hypothetical protein
reverse	22	12160	12032	129	42	hypothetical protein
reverse	23	12527	12225	303	100	hypothetical protein
reverse	24	12848	12600	249	82	small outer capsid protein
reverse	25	13409	12888	522	173	dCTPase
reverse	26	13754	13419	336	111	hypothetical protein
reverse	27	14084	13800	285	94	hypothetical protein
reverse	28	14428	14081	348	115	hypothetical protein
reverse	29	15437	14415	1023	340	DNA primase subunit
reverse	30	15950	15483	468	155	hypothetical protein

reverse	31	16256	15969	288	95	spackle periplasmic protein
reverse	32	17744	16314	1431	476	DNA primase-helicase subunit
reverse	33	18101	17757	345	114	head vertex assembly chaperone
reverse	34	19269	18103	1167	388	recA- like recombination protein
reverse	35	19547	19380	168	55	hypothetical protein
reverse	36	20140	19592	549	182	hypothetical protein
reverse	37	21039	20128	912	303	hypothetical protein
reverse	38	21737	21039	699	232	putative thymidylate synthase
reverse	39	22078	21734	345	114	hypothetical protein
reverse	40	23244	22075	1170	389	hypothetical protein
reverse	41	23884	23246	639	212	segB - putative endonuclease
reverse	42	24246	23881	366	121	cpb - capsule biosynthesis protein
reverse	43	25924	24230	1695	564	hypothetical protein
reverse	44	26106	25963	144	47	hypothetical protein
reverse	45	26732	26103	630	209	D-arabinose 5-phosphate isomerase
reverse	46	27142	26729	414	137	hypothetical protein
reverse	47	29911	27203	2709	902	DNA polymerase
reverse	48	30352	29990	363	120	regA - translational repressor protein
reverse	49	30929	30354	576	191	clamp loader small subunit
reverse	50	31876	30920	957	318	DNA polymerase clamp loader subunit
reverse	51	32637	31948	690	229	sliding clamp DNA polymerase accessory
reverse	52	33054	32683	372	123	rpbA - RNA polymerase binding protein
reverse	53	33254	33063	192	63	hypothetical protein
reverse	54	34894	33254	1641	546	endonuclease subunit
reverse	55	35952	34933	1020	339	recombination endonuclease subunit
reverse	56	36092	35949	144	47	hypothetical predicted membrane protein
reverse	57	36225	36157	69	22	hypothetical protein
reverse	58	36583	36398	186	61	hypothetical protein
reverse	59	36765	36568	198	65	hypothetical protein
reverse	60	37093	36743	351	116	hypothetical protein
reverse	61	37323	37096	228	75	hypothetical protein
reverse	62	37843	37307	537	178	sigma factor for late transcription
reverse	63	38173	37910	264	87	hypothetical protein
reverse	64	38403	38173	231	76	hypothetical protein
reverse	65	38753	38412	342	113	hypothetical protein
reverse	66	39074	38814	261	86	hypothetical protein
reverse	67	39456	39145	312	103	nrdH - glutaredoxin
reverse	68	39709	39437	273	90	hypothetical protein

reverse	69	39953	39711	243	80	hypothetical protein
reverse	70	40552	40079	474	157	nrdG - anaerobic NTP reductase, small subunit
reverse	71	42362	40539	1824	607	nrdD - anaerobic NTP reductase, large subunit
reverse	72	42832	42359	474	157	recombinase endonuclease VII
reverse	73	43169	42867	303	100	hypothetical protein
reverse	74	43620	43156	465	154	pinA - inhibitor of host Lon protease
reverse	75	44025	43723	303	100	conserved hypothetical protein
reverse	76	44212	44018	195	64	hypothetical protein
reverse	77	44365	44213	153	50	hypothetical protein
reverse	78	44662	44366	297	98	conserved hypothetical protein
reverse	79	44822	44655	168	55	hypothetical protein
reverse	80	44979	44815	165	54	hypothetical protein
reverse	81	45220	44945	276	91	thioredoxin
reverse	82	45437	45183	255	84	hypothetical protein
reverse	83	45795	45466	330	109	hypothetical protein
reverse	84	46300	45785	516	171	hypothetical protein
reverse	85	46527	46303	225	74	hypothetical protein
reverse	86	46760	46524	237	78	hypothetical protein
reverse	87	47793	46753	1041	346	hypothetical protein
reverse	88	48667	47894	774	257	hypothetical protein
reverse	89	48913	48725	189	62	hypothetical protein
reverse	90	49281	48913	369	122	hypothetical protein
reverse	91	49451	49281	171	56	hypothetical protein
reverse	92	49914	49444	471	156	hypothetical protein
reverse	93	50408	49914	495	164	hypothetical protein
reverse	94	50635	50405	231	76	hypothetical protein
reverse	95	51679	50699	981	326	hypothetical protein
reverse	96	52034	51798	237	78	hypothetical protein
reverse	97	52225	52031	195	64	hypothetical protein
reverse	98	53243	52230	1014	337	thioredoxin
reverse	99	53683	53240	444	147	hypothetical protein
reverse	100	53816	53652	165	54	hypothetical protein
reverse	101	54402	53884	519	172	hypothetical protein
reverse	102	54926	54441	486	161	hypothetical protein
reverse	103	55191	54928	264	87	hypothetical protein
reverse	104	55476	55249	228	75	hypothetical protein
reverse	105	56149	55478	672	223	hypothetical protein
reverse	106	56308	56180	129	42	hypothetical protein

reverse	107	56548	56441	108	35	hypothetical protein
reverse	108	57051	56605	447	148	hypothetical protein
reverse	109	57540	57154	387	128	hypothetical protein
reverse	110	57820	57527	294	97	rI - lysis inhibition regulator membrane protein
reverse	111	58042	57830	213	70	hypothetical protein
reverse	112	58488	58105	384	127	hypothetical protein
reverse	113	58720	58550	171	56	hypothetical protein
reverse	114	59297	58710	588	195	tk - thymidine kinase
reverse	115	59481	59299	183	60	tk - thymidine kinase
reverse	116	59948	59478	471	156	hypothetical protein
reverse	117	60301	59945	357	118	vs - valyl-tRNA synthetase modifier
reverse	118	60849	60301	549	182	hypothetical protein
reverse	119	61316	60861	456	151	regB - site-specific RNA endonuclease
reverse	120	61666	61376	291	96	hypothetical protein
reverse	121	61898	61656	243	80	hypothetical protein
reverse	122	62139	61873	267	88	hypothetical protein
reverse	123	62669	62136	534	177	hypothetical protein
reverse	124	63301	62885	417	138	denV - endonuclease V, N-glycosylase UV repair enzyme
reverse	125	63643	63356	288	95	hypothetical protein
reverse	126	64128	63643	486	161	lysozyme murein hydrolase
reverse	127	64635	64162	474	157	nudE - nudix hydrolase
reverse	128	64864	64628	237	78	hypothetical protein
reverse	129	65204	64875	330	109	hypothetical protein
reverse	130	65740	65201	540	179	hypothetical protein
reverse	131	66672	65740	933	310	hypothetical protein
reverse	132	66963	66739	225	74	hypothetical protein
reverse	133	67298	67005	294	97	hypothetical protein
reverse	134	67672	67298	375	124	hypothetical protein
reverse	135	68486	67665	822	273	hypothetical protein
reverse	136	69129	68524	606	201	hypothetical protein
reverse	137	69612	69217	396	131	hypothetical protein
reverse	138	70303	69698	606	201	hypothetical protein
reverse	139	70841	70497	345	114	hypothetical protein
reverse	140	71033	70949	85	NA	tRNA-Leu (TAA)
reverse	141	71115	71039	77	NA	tRNA-Arg (TCT)
reverse	142	71708	71259	450	149	hypothetical protein
reverse	143	71932	71708	225	74	hypothetical protein

reverse	144	72180	71929	252	83	hypothetical protein
reverse	145	72275	72200	76	NA	tRNA-Met (CAT)
reverse	146	72560	72333	228	114	hypothetical protein
reverse	147	72736	72662	75	NA	tRNA-Gly (TCC)
reverse	148	72811	72738	74	NA	tRNA-Trp (CCA)
reverse	149	72892	72818	74	NA	tRNA-Phe (GAA)
reverse	150	73055	72980	76	NA	tRNA-Ile (GAT)
reverse	151	73137	73060	78	NA	tRNA-Pro (TGG)
reverse	152	73384	73139	246	81	hypothetical protein
reverse	153	73503	73410	94	NA	tRNA-Ser (TGA)
reverse	154	73583	73508	76	NA	tRNA-His (GTG)
reverse	155	73755	73680	76	NA	tRNA-Gln (TTG)
reverse	156	74139	73762	378	125	hypothetical protein
reverse	157	74272	74196	77	NA	tRNA-Met (CAT)
reverse	158	74481	74398	84	NA	tRNA-Asn (GTT)
reverse	159	74565	74489	77	NA	tRNA-Lys (TTT)
reverse	160	74645	74571	75	NA	tRNA-Glu (TTC)
reverse	161	74725	74650	76	NA	tRNA-Asp (GTC)
reverse	162	74820	74732	89	NA	tRNA-Tyr (GTA)
reverse	163	75112	74924	189	62	hypothetical protein
reverse	164	75308	75105	204	67	hypothetical protein
reverse	165	75480	75295	186	61	hypothetical protein
reverse	166	75875	75543	333	110	hypothetical protein
reverse	167	76327	75872	456	151	hypothetical protein
reverse	168	77288	76596	693	230	dNMP kinase
reverse	169	77877	77290	588	195	tail completion and sheath stabilizer protein
reverse	170	78797	77967	831	276	DNA end protector protein
reverse	171	79246	78797	450	149	head completion protein
forward	172	79294	79869	576	191	baseplate wedge subunit
forward	173	79878	81596	1719	572	baseplate hub + tail lysozyme
forward	174	81596	82102	507	168	hypothetical protein
forward	175	82395	84347	1953	650	baseplate wedge subunit
forward	176	84344	87430	3087	1028	phage baseplate wedge initiator
forward	177	87430	88434	1005	334	phage baseplate wedge subunit
forward	178	88499	89371	873	290	baseplate wedge tail fiber connector
forward	179	89368	91197	1830	609	baseplate wedge subunit and tail pin
forward	180	91197	91880	684	227	baseplate wedge subunit and tail pin
forward	181	91880	93424	1545	514	short tail fiber protein

forward	182	93434	95188	1755	584	wac - fibritin neck whisker protein
forward	183	95248	96186	939	312	neck protein
forward	184	96188	96964	777	258	neck protein
forward	185	97007	97840	834	277	tail sheath stabilizer and completion protein
forward	186	97837	98337	501	166	terminase DNA packaging enzyme, small subunit
forward	187	98321	100156	1836	611	terminase DNA packaging enzyme, large subunit
forward	188	100192	102174	1983	660	hypothetical protein
forward	189	102275	102958	684	227	hypothetical protein
forward	190	102962	103453	492	163	tail tube protein
forward	191	103571	105139	1569	522	portal vertex protein of head
forward	192	105139	105390	252	83	prohead core protein
forward	193	105391	105816	426	141	prohead core protein
forward	194	105816	106463	648	215	prohead assembly (scaffolding) protein and protease
forward	195	106496	107311	816	271	prohead core scaffold protein
forward	196	107333	108892	1560	519	major capsid protein
forward	197	108981	110234	1254	417	phage capsid vertex
reverse	198	110813	110262	552	183	hypothetical protein
reverse	199	111820	110813	1008	335	rn1B - RNA ligase 2
reverse	200	112217	111924	294	97	conserved hypothetical protein
reverse	201	112441	112214	228	75	hypothetical protein
reverse	202	113058	112519	540	179	hoc - head outer capsid protein
reverse	203	113799	113089	711	236	inh - inhibitor of prohead protease
forward	204	114081	115358	1278	425	uvsW - RNA-DNA and DNA-DNA helicase, ATPase
forward	205	115355	115582	228	75	uvsW.1 - RNA-DNA and DNA-DNA helicase
reverse	206	115810	115646	165	54	hypothetical protein
reverse	207	116280	115846	435	144	uvsY - recombination repair and ssDNA binding protein
reverse	208	116732	116337	396	131	baseplate wedge subunit
reverse	209	117361	116732	630	209	baseplate hub subunit
forward	210	117413	118165	753	250	baseplate hub assembly catalyst
forward	211	118162	119382	1221	406	baseplate hub subunit
forward	212	119282	119821	540	179	baseplate hub distal subunit
forward	213	119805	121562	1758	585	base plate hub subunit, tail length determinator
forward	214	121571	122671	1101	366	phage baseplate tail tube cap
forward	215	122671	123636	966	321	baseplate subunit
reverse	216	124011	123670	342	113	hypothetical protein

reverse	217	126141	124039	2103	700	alt - RNA polymerase ADP-ribosylase
reverse	218	126389	126207	183	60	hypothetical protein
reverse	219	127882	126386	1497	498	DNA ligase
reverse	220	128157	127891	267	88	hypothetical protein
reverse	221	128899	128159	741	246	hypothetical protein
reverse	222	129053	128874	180	59	hypothetical protein
reverse	223	129244	129050	195	64	hypothetical protein
reverse	224	129589	129422	168	55	hypothetical protein
reverse	225	130035	129589	447	148	hypothetical protein
reverse	226	130425	130069	357	118	hypothetical protein
reverse	227	131008	130484	525	174	hypothetical protein
reverse	228	131486	131016	471	156	hypothetical protein
reverse	229	131891	131523	369	122	hypothetical protein
reverse	230	132173	131997	177	58	hypothetical protein
reverse	231	132582	132334	249	82	rIII - lysis inhibition accessory protein
reverse	232	133004	132681	324	107	head assembly co-chaperone
reverse	233	133385	133062	324	107	hypothetical protein
reverse	234	133949	133386	564	187	cd - dCMP deaminase
reverse	235	134290	133949	342	113	hypothetical protein
reverse	236	135287	134271	1017	338	phospho-2-dehydro-3-deoxyheptonate aldolase
reverse	237	135517	135287	231	76	hypothetical protein
reverse	238	135785	135519	267	88	hypothetical protein
reverse	239	136684	135785	900	299	polynucleotide kinase
reverse	240	136899	136681	219	72	hypothetical protein
reverse	241	137378	136896	483	160	hypothetical protein
reverse	242	137817	137518	300	99	phage outer membrane lipoprotein
reverse	243	138170	137814	357	118	putative membrane protein
reverse	244	138658	138158	501	166	alc - inhibitor of host transcription
reverse	245	139891	138719	1173	390	rnlA - RNA ligase 1 and tail fiber attachment
reverse	246	140301	139888	414	137	denA - endonuclease II
reverse	247	141512	140334	1179	392	nrdB - aerobic NDP reductase, small subunit
reverse	248	141706	141509	198	65	hypothetical protein
reverse	249	144000	141745	2256	751	nrdA - aerobic NDP reductase, large subunit
reverse	250	144296	143991	306	101	hypothetical protein
reverse	251	145153	144293	861	286	thyA - dTMP thymidylate synthase
reverse	252	145524	145150	375	124	hypothetical protein
reverse	253	145846	145505	342	113	hypothetical protein
reverse	254	146429	145827	603	200	dihydrofolate reductase

reverse	255	146650	146426	225	74	hypothetical protein
reverse	256	147020	146742	279	92	hypothetical protein
reverse	257	147432	147076	357	118	hypothetical protein
reverse	258	147750	147517	234	77	hypothetical protein
reverse	259	147971	147747	225	74	hypothetical protein
reverse	260	148287	148036	252	83	hypothetical protein
reverse	261	149317	148430	888	295	single-stranded DNA binding protein
reverse	262	150053	149400	654	217	loader of gp41 DNA helicase
reverse	263	150636	150028	609	202	segD - homing endonuclease
reverse	264	150944	150633	312	103	late promoter transcription accessory protein
reverse	265	151206	150919	288	95	dsbA - dsDNA binding protein, late transcription
reverse	266	152155	151214	942	313	rnh - RNaseH ribonuclease
forward	267	152261	156031	3771	1256	long tail fiber, proximal subunit
forward	268	156040	157191	1152	383	hinge connector of long tail fiber, proximal connector
forward	269	157253	157912	660	219	hinge connector
forward	270	157921	160665	2745	914	long tail fiber, distal subunit
forward	271	160697	161212	516	171	distal long tail fiber assembly catalyst
forward	272	161393	161899	507	168	receptor-recognizing protein
forward	273	161927	162583	657	218	holin lysis mediator
reverse	274	162870	162601	270	89	asiA - anti-sigma 70 protein
reverse	275	163262	162924	339	112	hypothetical protein
reverse	276	163377	163246	132	43	hypothetical protein
reverse	277	163570	163352	219	72	hypothetical protein
reverse	278	163847	163542	306	101	hypothetical protein
reverse	279	164281	163847	435	144	hypothetical protein
reverse	280	164622	164278	345	114	hypothetical protein
reverse	281	165272	164637	636	211	motA - transcriptional regulator of middle promoters
reverse	282	166040	165369	672	223	hypothetical protein
reverse	283	166349	166209	141	46	hypothetical protein
reverse	284	167713	166349	1365	454	topoisomerase II medium subunit
reverse	285	167873	167718	156	51	hypothetical protein
reverse	286	168623	168195	429	142	ndd - nucleoid disruption protein
reverse	287	168909	168709	201	66	hypothetical protein
reverse	288	169039	168917	123	40	hypothetical protein
reverse	289	169724	169077	648	215	endonuclease IV
reverse	290	169947	169645	303	100	hypothetical protein

reverse	291	170178	169978	201	66	hypothetical protein
reverse	292	171151	170222	930	309	rIIB - protector from prophage-induced early lysis

Appendix II - Genome annotation of ΦCBH189

Strand	Gene number	Start position	Stop position	Gene length (b.p.)	Product length (a.a.)	Predicted gene product
reverse	1	2169	1	2169	722	rIIA - protector from phage-induced early lysis
reverse	2	2383	2180	204	67	hypothetical protein
reverse	3	2741	2370	372	123	hypothetical protein
reverse	4	2951	2772	180	59	hypothetical protein
reverse	5	4806	2968	1839	612	DNA topoisomerase II large subunit
reverse	6	5340	5092	249	82	hypothetical protein
reverse	7	5519	5340	180	59	conserved putative regulatory protein, fmdB family
reverse	8	5844	5506	339	112	hypothetical protein
reverse	9	6268	5834	435	144	hypothetical protein
reverse	10	6624	6268	357	118	hypothetical protein
reverse	11	6842	6624	219	72	cef - modifier of suppressor tRNAs
reverse	12	7328	6930	399	132	hypothetical protein
reverse	13	8022	7339	684	227	exonuclease A
reverse	14	8248	8015	234	77	hypothetical protein
reverse	15	9581	8256	1326	441	RecD-like DNA helicase
reverse	16	9877	9581	297	98	hypothetical protein
reverse	17	10572	9877	696	231	putative anti-sigma factor
reverse	18	11267	10659	609	202	ADP-ribosylase
reverse	19	11505	11323	183	60	hypothetical protein
reverse	20	11850	11602	249	82	hypothetical protein
reverse	21	12418	12170	249	82	small outer capsid protein
reverse	22	12979	12458	522	173	dCTPase
reverse	23	13324	12989	336	111	hypothetical protein
reverse	24	13654	13370	285	94	hypothetical protein
reverse	25	13784	13665	120	39	hypothetical protein
reverse	26	14793	13771	1023	340	DNA primase subunit
reverse	27	15306	14839	468	155	hypothetical protein
reverse	28	15612	15325	288	95	spackle periplasmic protein
reverse	29	17100	15670	1431	476	DNA primase-helicase subunit

reverse	30	17457	17113	345	114	head vertex assembly chaperone
reverse	31	18625	17459	1167	388	recA-like recombination protein
reverse	32	19496	18948	549	182	hypothetical protein
reverse	33	20395	19484	912	303	hypothetical protein
reverse	34	21093	20395	699	232	putative thymidylate synthase
reverse	35	21458	21090	369	122	hypothetical protein
reverse	36	22627	21458	1170	389	peptidase
reverse	37	23267	22629	639	212	homing endonuclease
reverse	38	23629	23264	366	121	hypothetical protein
reverse	39	25307	23613	1695	564	hypothetical protein
reverse	40	25489	25346	144	47	hypothetical protein
reverse	41	26115	25486	630	209	D-arabinose 5-phosphate isomerase
reverse	42	26531	26112	420	139	hypothetical protein
reverse	43	29294	26586	2709	902	DNA polymerase
reverse	44	29735	29373	363	120	regA - translational repressor protein
reverse	45	30312	29737	576	191	clamp loader small subunit
reverse	46	31274	30303	972	323	DNA polymerase clamp loader subunit
reverse	47	32020	31331	690	229	sliding clamp DNA polymerase accessory
reverse	48	32437	32066	372	123	rpbA - RNA polymerase binding protein
reverse	49	32637	32446	192	63	hypothetical protein
reverse	50	34319	32637	1683	560	recombination endonuclease subunit
reverse	51	34531	34316	216	71	hypothetical protein
reverse	52	35543	34524	1020	339	recombination endonuclease subunit
reverse	53	35971	35816	156	51	hypothetical protein
reverse	54	36174	35989	186	61	hypothetical protein
reverse	55	36356	36159	198	65	hypothetical protein
reverse	56	36684	36334	351	116	hypothetical protein
reverse	57	36914	36687	228	75	hypothetical protein
reverse	58	37434	36898	537	178	sigma factor for late transcription
reverse	59	37764	37501	264	87	hypothetical protein
reverse	60	37994	37764	231	76	hypothetical protein
reverse	61	38380	38003	378	125	hypothetical protein
reverse	62	38665	38405	261	86	hypothetical protein
reverse	63	39038	38736	303	100	glutaredoxin
reverse	64	39300	39028	273	90	hypothetical protein
reverse	65	39544	39302	243	80	hypothetical protein
reverse	66	40143	39670	474	157	anaerobic NTP reductase small subunit

reverse	67	41953	40130	1824	607	anaerobic NTP reductase large subunit
reverse	68	42423	41950	474	157	recombinase endonuclease VII
reverse	69	42760	42458	303	100	hypothetical protein
reverse	70	43184	42747	438	145	inhibitor of host Lon protease
reverse	71	43329	43165	165	54	hypothetical protein
reverse	72	43616	43314	303	100	hypothetical protein
reverse	73	43803	43609	195	64	hypothetical protein
reverse	74	43956	43804	153	50	hypothetical protein
reverse	75	44253	43957	297	98	hypothetical protein
reverse	76	44413	44246	168	55	hypothetical protein
reverse	77	44811	44536	276	91	thioredoxin
reverse	78	45076	44774	303	100	hypothetical protein
reverse	79	45386	45057	330	109	hypothetical protein
reverse	80	45891	45376	516	171	hypothetical protein
reverse	81	46118	45894	225	74	hypothetical protein
reverse	82	46351	46115	237	78	hypothetical protein
reverse	83	47384	46344	1041	346	hypothetical protein
reverse	84	48288	47485	804	267	hypothetical protein
reverse	85	48504	48316	189	62	hypothetical protein
reverse	86	48872	48504	369	122	hypothetical protein
reverse	87	49042	48872	171	56	hypothetical protein
reverse	88	49505	49035	471	156	hypothetical protein
reverse	89	49999	49505	495	164	hypothetical protein
reverse	90	50262	49996	267	88	hypothetical protein
reverse	91	51339	50290	1050	349	thioredoxin
reverse	92	51628	51392	237	78	hypothetical protein
reverse	93	52641	51625	1017	338	thioredoxin
reverse	94	53060	52638	423	140	hypothetical protein
reverse	95	53289	53050	240	79	hypothetical protein
reverse	96	53794	53282	513	170	hypothetical protein
reverse	97	54318	53833	486	161	hypothetical protein
reverse	98	54542	54318	225	74	hypothetical protein
reverse	99	55516	54830	687	228	hypothetical protein
reverse	100	55793	55527	267	88	putative C4-type zinc finger protein
reverse	101	56413	55952	462	153	hypothetical protein
reverse	102	56903	56517	387	128	hypothetical protein
reverse	103	57183	56890	294	97	lysis inhibitor regulator

reverse	104	57405	57193	213	70	hypothetical protein
reverse	105	57851	57468	384	127	hypothetical protein
reverse	106	58083	57913	171	56	hypothetical protein
reverse	107	58660	58073	588	195	thymidine kinase
reverse	108	58871	58662	210	69	hypothetical protein
reverse	109	59047	58871	177	58	hypothetical protein
reverse	110	59510	59040	471	156	hypothetical protein
reverse	111	59866	59507	360	119	valyl-tRNA synthetase modifier
reverse	112	60408	59863	546	181	hypothetical protein
reverse	113	60875	60420	456	151	site-specific RNA endonuclease
reverse	114	61212	60937	276	91	hypothetical protein
reverse	115	61444	61202	243	80	hypothetical protein
reverse	116	61685	61419	267	88	hypothetical protein
reverse	117	62350	61682	669	222	hypothetical protein
reverse	118	62696	62397	300	99	putative internal head protein
reverse	119	63157	62741	417	138	pyrimidine dimer DNA glycosylase
reverse	120	63499	63212	288	95	hypothetical protein
reverse	121	63984	63499	486	161	T4-like lysozyme
reverse	122	64491	64018	474	157	hypothetical protein
reverse	123	64720	64484	237	78	hypothetical protein
reverse	124	65060	64731	330	109	hypothetical protein
reverse	125	65371	65057	315	104	hypothetical protein
reverse	126	65840	65361	480	159	hypothetical protein
reverse	127	66842	65910	933	310	hypothetical protein
reverse	128	67175	66909	267	88	hypothetical protein
reverse	129	67468	67175	294	97	hypothetical protein
reverse	130	67842	67468	375	124	hypothetical protein
reverse	131	68656	67835	822	273	hypothetical protein
reverse	132	69299	68694	606	201	hypothetical protein
reverse	133	69584	69384	201	66	hypothetical protein
reverse	134	70051	69656	396	131	hypothetical protein
reverse	135	70699	70136	564	187	hypothetical protein
reverse	136	71188	70799	390	129	hypothetical protein
reverse	137	71265	71181	85	NA	tRNA-Leu (TAA)
reverse	138	71346	71271	76	NA	tRNA-Arg (TCT)
reverse	139	71939	71490	450	149	hypothetical protein
reverse	140	72163	71939	225	74	hypothetical protein

reverse	141	72495	72160	336	111	hypothetical protein
reverse	142	72509	72434	76	NA	tRNA-Met (CAT)
reverse	143	72956	72567	390	129	hypothetical protein
reverse	144	72970	72896	75	24	tRNA-Gly (TCC)
reverse	145	73051	72978	74	NA	tRNA-Trp (CCA)
reverse	146	73279	73205	75	24	tRNA-Phe (GAA)
reverse	147	73441	73366	76	NA	tRNA-Ile (GAT)
reverse	148	73523	73446	78	25	tRNA-Pro (TGG)
reverse	149	73770	73525	246	81	hypothetical protein
reverse	150	73887	73796	92	NA	tRNA-Ser (TGA)
reverse	151	73967	73892	76	NA	tRNA-His (GTG)
reverse	152	74139	74064	76	NA	tRNA-Gln (TTG)
reverse	153	74555	74148	408	135	adhesin
reverse	154	74657	74581	77	NA	tRNA-Met (CAT)
reverse	155	74860	74784	77	NA	tRNA-Asn (GTT)
reverse	156	75088	75013	76	NA	tRNA-Asp (GTC)
reverse	157	75273	75079	195	64	hypothetical protein
reverse	158	75183	75095	89	NA	tRNA-Tyr (GTA)
reverse	159	75485	75291	195	64	hypothetical protein
reverse	160	75654	75475	180	59	hypothetical protein
reverse	161	75851	75651	201	66	hypothetical protein
reverse	162	76246	75914	333	110	hypothetical protein
reverse	163	76698	76243	456	151	hypothetical protein
reverse	164	76967	76698	270	89	chaperone for long tail fiber formation
reverse	165	77659	76967	693	230	dNMP kinase
reverse	166	78248	77661	588	195	tail completion and sheath stabilizer protein
reverse	167	79168	78338	831	276	DNA end protector protein
reverse	168	79617	79168	450	149	head completion protein
forward	169	79665	80240	576	191	baseplate wedge subunit
forward	170	80240	81967	1728	575	baseplate hub subunit and tail lysozyme
forward	171	81967	82473	507	168	hypothetical protein
forward	172	82473	82766	294	97	hypothetical protein
forward	173	82766	84718	1953	650	baseplate wedge subunit
forward	174	84715	87801	3087	1028	baseplate wedge initiator
forward	175	87801	88805	1005	334	baseplate wedge subunit
forward	176	88870	89742	873	290	baseplate wedge tail fiber connector
forward	177	89739	91568	1830	609	baseplate wedge subunit and tail pin

forward	178	91568	92251	684	227	baseplate wedge subunit and tail pin
forward	179	92251	93795	1545	514	short tail fiber protein
forward	180	93805	95559	1755	584	fibritin neck whisker protein
forward	181	95619	96557	939	312	neck protein
forward	182	96559	97335	777	258	tail sheath stabilizer and completion protein
forward	183	97426	98211	786	261	terminase DNA packaging enzyme small subunit
forward	184	98208	98708	501	166	terminase DNA packaging enzyme large subunit
forward	185	98692	100527	1836	611	tail sheath protein
forward	186	100563	102545	1983	660	hypothetical protein
forward	187	102646	103329	684	227	tail tube protein
forward	188	103333	103824	492	163	portal vertex protein of head
forward	189	103942	105510	1569	522	prohead core protein
forward	190	105510	105779	270	89	prohead core protein
forward	191	105780	106205	426	141	prohead assembly (scaffolding) protein and protease
forward	192	106205	106852	648	215	prohead core scaffold protein
forward	193	106885	107700	816	271	major capsid protein
forward	194	107722	109281	1560	519	capsid vertex protein
forward	195	109371	110624	1254	417	hypothetical protein
reverse	196	111203	110652	552	183	RNA ligase 2
reverse	197	112210	111203	1008	335	hypothetical protein
reverse	198	112577	112314	264	87	hypothetical protein
reverse	199	112807	112574	234	77	hypothetical protein
reverse	200	113424	112885	540	179	head outer capsid protein
reverse	201	114166	113456	711	236	inhibitor of prohead protease
forward	202	114223	115725	1503	500	RNA-DNA and DNA-DNA helicase, ATPase
forward	203	115722	115949	228	75	RNA-DNA and DNA-DNA helicase
reverse	204	116647	116213	435	144	recombination repair and ssDNA binding protein
reverse	205	117099	116704	396	131	baseplate wedge subunit
reverse	206	117728	117099	630	209	baseplate hub subunit
forward	207	117780	118532	753	250	baseplate hub assembly catalyst
forward	208	118529	119749	1221	406	baseplate hub subunit
forward	209	119784	120188	405	134	baseplate hub distal subunit
forward	210	120172	121917	1746	581	baseplate hub subunit tail length determinator
forward	211	121926	123026	1101	366	baseplate tail tube cap
forward	212	123026	123991	966	321	gp54 baseplate tail tube initiator
reverse	213	124365	124024	342	113	hypothetical protein

reverse	214	126495	124393	2103	700	RNA polymerase ADP-ribosylase
reverse	215	126743	126561	183	60	hypothetical protein
reverse	216	128236	126740	1497	498	DNA ligase
reverse	217	128523	128245	279	92	hypothetical protein
reverse	218	129253	128513	741	246	hypothetical protein
reverse	219	129410	129228	183	60	hypothetical protein
reverse	220	129598	129407	192	63	hypothetical protein
reverse	221	129779	129573	207	68	hypothetical protein
reverse	222	129943	129776	168	55	hypothetical protein
reverse	223	130395	129943	453	150	hypothetical protein
reverse	224	130801	130436	366	121	hypothetical protein
reverse	225	131384	130860	525	174	hypothetical protein
reverse	226	131862	131392	471	156	hypothetical protein
reverse	227	132253	131900	354	117	hypothetical protein
reverse	228	132529	132359	171	56	hypothetical protein
reverse	229	132892	132644	249	82	rIII lysis inhibition accessory protein
reverse	230	133314	132991	324	107	head assembly co-chaperone
reverse	231	133695	133372	324	107	tail fiber protein
reverse	232	134259	133696	564	187	dCMP deaminase
reverse	233	134600	134259	342	113	hypothetical protein
reverse	234	135597	134581	1017	338	phospho-2-dehydro-deoxyheptonate aldolase
reverse	235	135827	135597	231	76	hypothetical protein
reverse	236	136095	135829	267	88	hypothetical protein
reverse	237	136295	136092	204	67	hypothetical protein
reverse	238	137040	136327	714	237	Seg-like homing endonuclease
reverse	239	137995	137096	900	299	polynucleotide kinase
reverse	240	138231	137992	240	79	hypothetical protein
reverse	241	138796	138311	486	161	hypothetical protein
reverse	242	139235	138936	300	99	outer membrane lipoprotein
reverse	243	139588	139232	357	118	putative membrane protein
reverse	244	140070	139576	495	164	inhibitor of host transcription
reverse	245	141303	140131	1173	390	RNA ligase 1 and tail fiber attachment catalyst
reverse	246	141713	141300	414	137	DenA endonuclease II
reverse	247	143005	141746	1260	419	NrdB aerobic NDP reductase, small subunit
reverse	248	145217	142962	2256	751	NrdA aerobic NDP reductase, large subunit
reverse	249	145513	145208	306	101	hypothetical protein
reverse	250	145698	145510	189	62	dTMP (thymidylate) synthase

reverse	251	146555	145695	861	286	hypothetical protein
reverse	252	146926	146555	372	123	hypothetical protein
reverse	253	147248	146907	342	113	dihydrofolate reductase
reverse	254	147831	147229	603	200	hypothetical protein
reverse	255	148097	147831	267	88	hypothetical protein
reverse	256	148482	148168	315	104	hypothetical protein
reverse	257	148924	148502	423	140	hypothetical protein
reverse	258	149177	148944	234	77	hypothetical protein
reverse	259	149398	149174	225	74	hypothetical protein
reverse	260	149798	149460	339	112	hypothetical protein
reverse	261	150866	149979	888	295	single-stranded DNA binding protein
reverse	262	151603	150950	654	217	loader of DNA helicase
reverse	263	151908	151600	309	102	late promoter transcription accessory protein
reverse	264	152188	151883	306	101	dsDNA binding protein
reverse	265	153119	152178	942	313	RNase H
forward	266	153225	156995	3771	1256	long tail fiber proximal subunit
forward	267	157004	158155	1152	383	hinge connector of long tail fiber, proximal connector
forward	268	158217	158876	660	219	hinge connector of long tail fiber, distal connector
forward	269	158885	161638	2754	917	long tail fiber distal subunit
forward	270	161670	162185	516	171	distal long tail fiber assembly catalyst
forward	271	162355	162999	645	214	hypothetical protein
forward	272	163027	163683	657	218	holin lysis mediator
reverse	273	163970	163701	270	89	asiA-anti-sigma 70 protein
reverse	274	164362	164024	339	112	hypothetical protein
reverse	275	164477	164346	132	43	hypothetical protein
reverse	276	164947	164651	297	98	anti-restriction nuclease
reverse	277	165381	164947	435	144	anti-restriction nuclease
reverse	278	165725	165378	348	115	anti-restriction nuclease
reverse	279	166372	165737	636	211	motA-transcriptional regulator of middle promoters
reverse	280	167182	166466	717	238	hypothetical protein
forward	281	167181	167453	273	90	hypothetical protein
reverse	282	168810	167446	1365	454	topoisomerase II medium subunit
reverse	283	168940	168815	126	41	hypothetical protein
forward	284	168956	169126	171	56	hypothetical protein
reverse	285	169292	169119	174	57	hypothetical protein
reverse	286	169729	169292	438	145	nucleoid disruption protein

reverse	287	169988	169797	192	63	hypothetical protein
reverse	288	170118	169996	123	40	hypothetical protein
reverse	289	170797	170156	642	213	denB endonuclease IV
reverse	290	171251	171051	201	66	hypothetical protein
reverse	291	172224	171295	930	309	rIIB protector from prophage-induced early lysis

Appendix III - Phenotypic comparison of *S*39006-wt and *S*39006-R4

Secondary metabolite production

To investigate whether OmpW is involved in the secretion of the Carbapenem, a β-lactam antibiotic with broad-spectrum activity (102), overnight cultures of *S*39006-wt and *S*39006-R4 were spotted side by side on a top lawn of *E.coli* ESS to compare their ability to produce a halo from killing ESS. As shown in Figure S1a, no obvious different was observed between *S*39006-wt and *S*39006-R4, indicating that the lack of OmpW does not affect ability to produce or transport the Carbapenem.

The role of OmpW in the production or excretion of prodigiosin, the red pigment molecule with antimalarial, antifungal and antibiotic function, was also investigated by comparing the amount of prodigiosin secreted by *S*39006-wt and *S*39006-R4 in exponential phase (116). As shown in Figure S1b, no significant difference in prodigiosin secretion was found between *S*39006-wt and *S*39006-R4, indicating that OmpW is not involved in prodigiosin production or secretion.

Furthermore, the effect of OmpW on the production of N-butanoyl-L-homoserine lactone (BHL) was also investigated. BHL is a quorum sensing molecule that mediates the control of secondary metabolite production in *S*39006. However, OmpW was not involved in BHL production (data not shown).

Motility

*S*39006 is able to facilitate its own motility through swimming, swarming and flotation. To investigate whether OmpW deficiency affects any of the motility mechanisms, *S*39006-wt and *S*39006-R4 were compared with regard to their ability to swim on Tryptone agar, swarm on

Eiken agar and float in LB. However, no significant differences were observed in any of the assays (data not shown), indicating that OmpW is not involved in motility of *S*39006.

Stress response

The role of OmpW in stress response was investigated by comparing the growth of *S*39006-wt and *S*39006-R4 on solid agar containing 2.5% ethanol, 3% H_2O_2, 0.1% SDS and 2.5% NaCl, respectively. As shown in Figure S1c, no obvious difference was observed under these stress conditions between *S*39006-wt and *S*39006-R4, indicating that OmpW is not involved in the response to the stress conditions tested.

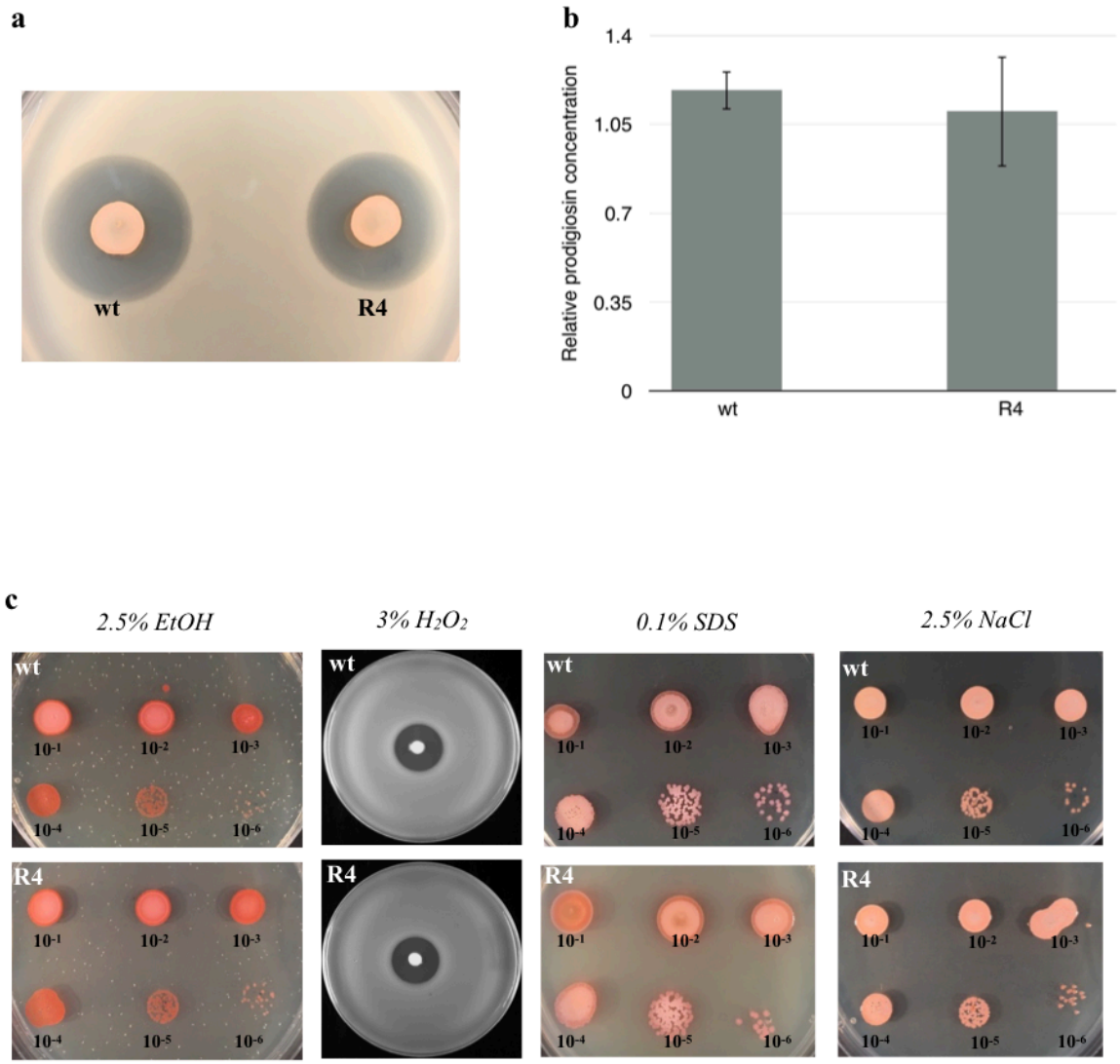


Figure S1: Secondary metabolite secretion and stress response in *S*39006-wt and *S*39006-R4. (a) Comparison of Carbapenem production in *S*39006-wt and *S*39006-R4. The halo surrounding spotted bacterial culture is formed from Carbapenem inhibition of *E.coli* ESS growth in the top lawn, therefore the size of the halo is reflective of Carbapenem concentration. The result shown is representative of triplicates. (b) Comparison of relative prodigiosin concentration in *S*39006-wt and *S*39006-R4. Error bars represent standard deviation from triplicates. (c) Comparison of *S*39006-wt and *S*39006-R4 growth response to 2.5% ethanol, 3% H_2O_2, 0.1% SDS and 2.5% NaCl. For ethanol, SDS and NaCl, overnight cultures of cells were serially diluted and spotted on LBA containing cognate stress agents. For H_2O_2, filter paper containing 3% H_2O_2 was placed on a top lawn containing *S*39006-wt or *S*39006-R4. The smaller the size of the clearing around the filter paper, the higher the level of bacterial tolerance against H_2O_2. The result shown is representative of triplicates.

Antibiotic resistance

Previous research has shown that OmpW is related to antibiotic resistance in other species (168). To investigate whether the same is true for *S*39006, an assay was performed comparing the zone of clearing formed by different antibiotic disks on the top lawns of *S*39006-wt and *S*39006-R4. The inhibition ratio was calculated by dividing the clearance diameter formed on *S*39006-wt by the clearance diameter formed on *S*39006-R4. As shown in Figure S2a, of the 25 antibiotics tested, only three of them had no growth inhibition effect on either *S*39006-wt or *S*39006-R4. However, for those active against *S*39006-wt, no difference in action was observed for the OmpW-mutant as all of their inhibition ratio are close to 1. This suggests that OmpW is not involved in the cell's susceptibility to the antibiotics tested, either as an influx route or efflux pump.

Virulence

Previous studies have also shown that OmpW is involved in antigenic response from the host during bacterial infection (170). This prompted the interest to test whether OmpW is involved in *S*39006 virulence during infection of its hosts, such as *C. elegans* and potato. In a *C.elegans* virulence assay, *S*39006-wt and *S*39006-R4 were used to feed *C. elegans* DH26 and virulence was reflected by the speed of DH26 dying. As shown in Figure S2b, when DH26 was fed with *E.coli* OP50 (negative control), the number of live worms decreased gradually until after ~ 14 days when no surviving worm was left. In comparison, worms fed on *S*39006-wt and the OmpW-mutant died rapidly with no worms alive after 4-5 days. However, the rate of decrease only had marginal differences with or without OmpW, suggesting that OmpW is not related to *S*39006 virulence in *C. elegans* DH26.

Similarly, in a potato virulence assay, *S*39006-wt and *S*39006-R4 were compared for their ability to cause soft rot. The same amount (1×10^6 cfu) of *S*39006-wt and the OmpW-mutant were inoculated to the same potato and the resulting soft rot was weighed every 24 hours. As shown in Figure S2c, no obvious difference in rot weight was observed during the infection process between *S*39006-wt and *S*39006-R4, indicating that OmpW is not involved in *S*39006 virulence against potato.

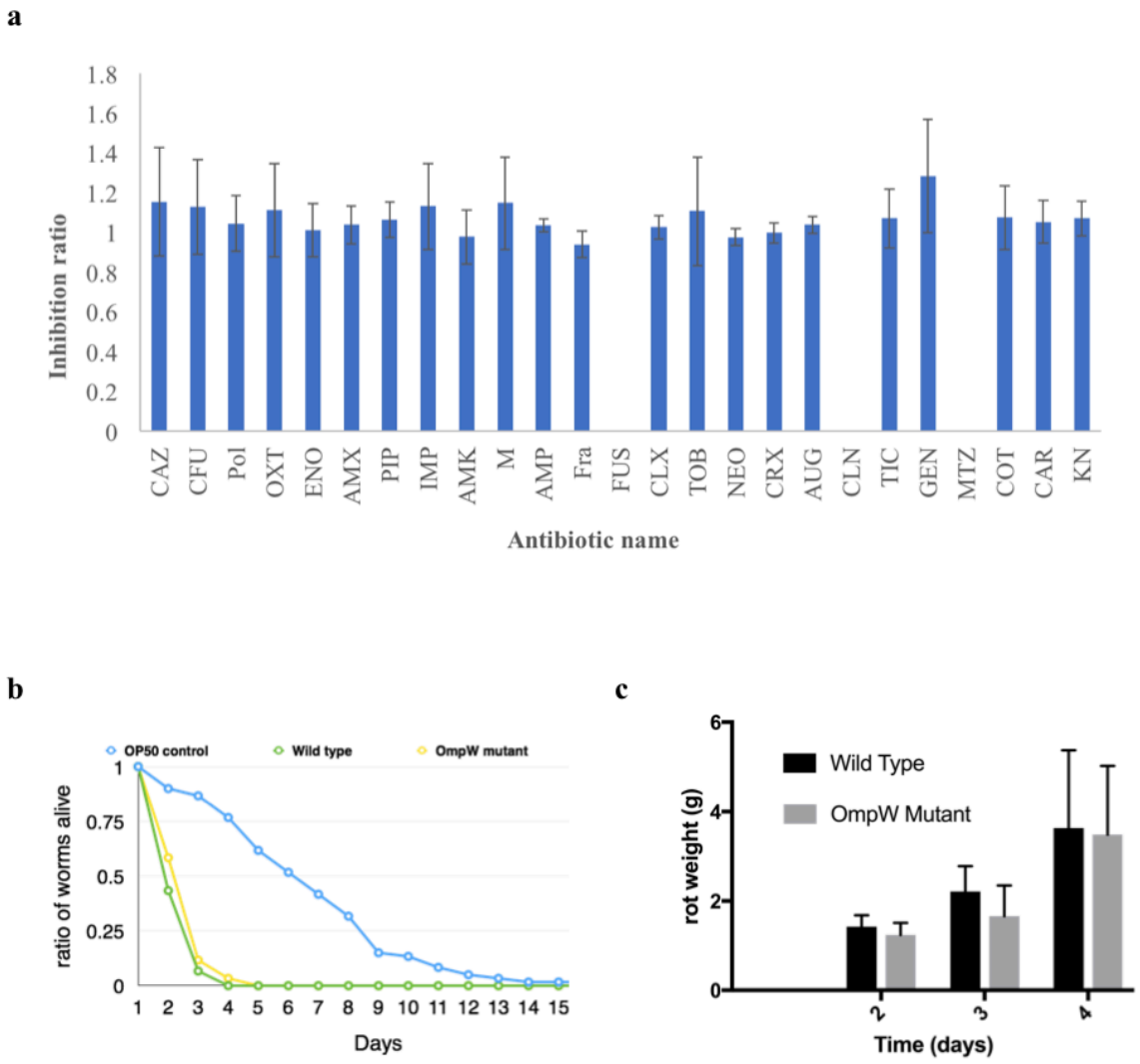


Figure S2: Antibiotic susceptibility and virulence of *S*39006-wt and *S*39006-R4.

(a) Antibiotic susceptibility of *S*39006-wt versus *S*39006-R4. The inhibition ratio is calculated by dividing the diameter of clearance formed around an antibiotic disk on *S*39006-wt by the diameter of clearance formed on *S*39006-R4. Where inhibition ratio is "0", it means the antibiotic was not able to inhibit growth on either *S*39006-wt or *S*39006-R4. Error bars represent standard deviation from triplicates. Abbreviation and concentration of antibiotics: CAZ – Ceftazidime 30µg/µl; CFU – Ceftiofur 20µg/µl; Pol – Polymyxin 100 IU; OXT – Oxytetracycline 30µg/µl; ENO – Enrofloxacin 5µg/µl; AMX – Amoxycillin 30µg/µl; PIP – Pipericillin 75µg/µl; IMP – Imipenem 10µg/µl; AMK – Amikacin 30µg/µl; M – Marbofloxacin 5µg/µl; AMP – Ampicillin 10µg/µl; Fra – Framycetin 100µg/µl; FUS - Fusidic Acid 10µg/µl; CLX – Cephalexin 30µg/µl; TOB – Tobramycin 10µg/µl; NEO – Neomycin 10µg/µl; CRX – Cefuroxime 30µg/µl; AUG – Augmentin 30µg/µl; CLN – Clindamycin 2µg/µl; TIC – Ticarcillin 75µg/µl; GEN – Gentamicin 10µg/µl;

MTZ – Metronidazole 5µg/µl; COT – Cotrimoxazole 25µg/µl; CAR – Carbenicillin 100µg/µl; KN – Kanamycin 5µg/µl

(b) Virulence comparison of *S*39006-wt and *S*39006-R4 infecting *C. elegans* DH26. The ratio of worms alive is calculated by dividing the number of live worms each day by the number of live worms on day 1. *E.coli* OP50 is used as negative control. The result shown is representative of triplicates. (c) Virulence comparison of *S*39006-wt and *S*39006-R4 infecting potato. The weight of soft rot was compared at day 2-4 post inoculation to reflect the level of virulence. Error bars represent standard deviation from triplicates.

Appendix IV - List of Publications

frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 May 2017
doi: 10.3389/fmicb.2017.01006

Environmental T4-Family Bacteriophages Evolve to Escape Abortive Infection via Multiple Routes in a Bacterial Host Employing "Altruistic Suicide" through Type III Toxin-Antitoxin Systems

*Bihe Chen, Chidiebere Akusobi †, Xinzhe Fang and George P. C. Salmond **

Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

OPEN ACCESS

Edited by:
William Michael McShan, University of Oklahoma Health Sciences Center, United States

Reviewed by:
Robert Czajkowski, University of Gdańsk, Poland
Scott Van Nguyen, Agricultural Research Service (USDA), United States

***Correspondence:**
George P. C. Salmond
gpcs2@cam.ac.uk

†Present Address:
Chidiebere Akusobi, Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Boston, MA, USA

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *23 March 2017*
Accepted: *19 May 2017*
Published: *31 May 2017*

Citation:
Chen B, Akusobi C, Fang X and Salmond GPC (2017) Environmental T4-Family Bacteriophages Evolve to Escape Abortive Infection via Multiple Routes in a Bacterial Host Employing "Altruistic Suicide" through Type III Toxin-Antitoxin Systems. Front. Microbiol. 8:1006. doi: 10.3389/fmicb.2017.01006

Abortive infection is an anti-phage mechanism employed by a bacterium to initiate its own death upon phage infection. This reduces, or eliminates, production of viral progeny and protects clonal siblings in the bacterial population by an act akin to an "altruistic suicide." Abortive infection can be mediated by a Type III toxin-antitoxin system called ToxIN$_{Pa}$ consisting of an endoribonuclease toxin and RNA antitoxin. ToxIN$_{Pa}$ is a heterohexameric quaternary complex in which pseudoknotted RNA inhibits the toxicity of the toxin until infection by certain phages causes destabilization of ToxIN$_{Pa}$, leading to bacteriostasis and, eventually, lethality. However, it is still unknown why only certain phages are able to activate ToxIN$_{Pa}$. To try to address this issue we first introduced ToxIN$_{Pa}$ into the Gram-negative enterobacterium, *Serratia* sp. ATCC 39006 (*S* 39006) and then isolated new environmental *S* 39006 phages that were scored for activation of ToxIN$_{Pa}$ and abortive infection capacity. We isolated three T4-like phages from a sewage treatment outflow point into the River Cam, each phage being isolated at least a year apart. These phages were susceptible to ToxIN$_{Pa}$-mediated abortive infection but produced spontaneous "escape" mutants that were insensitive to ToxIN$_{Pa}$. Analysis of these resistant mutants revealed three different routes of escaping ToxIN$_{Pa}$, namely by mutating *asiA* (the product of which is a phage transcriptional co-activator); by mutating a conserved, yet functionally unknown, *orf84*; or by deleting a 6.5–10 kb region of the phage genome. Analysis of these evolved escape mutants may help uncover the nature of the corresponding phage product(s) involved in activation of ToxIN$_{Pa}$.

Keywords: abortive infection, toxin-antitoxin, bacteriophage, *Serratia*, T4-family phage

INTRODUCTION

Bacteria are susceptible to viral (bacteriophage) predation but have evolved several strategies to resist viral infection. One strategy is abortive infection (Abi) in which an infected bacterial cell dies precociously and thereby concomitantly blocks the production of mature phage progeny (Chopin et al., 2005). This protects clonal siblings in the bacterial population and therefore is akin

to an "altruistic suicide." Abi can be mediated through toxin-antitoxin (TA) systems, which are widespread in prokaryotes. Genetically, TA systems are usually composed of two genes transcribed from a single promoter. The upstream gene encodes an antitoxin that neutralizes the toxin product of the downstream gene. Functionally, TA systems impact on several biological processes such as the formation of persister cells, responses to environmental stress, plasmid stabilization, and phage exclusion through abortive infection (Gerdes et al., 2005).

Some Type III TA systems are bifunctional in that they also confer an Abi phenotype on their bacterial hosts. These TA systems are comprised of a proteinaceous toxin and an RNA antitoxin (Fineran et al., 2009). The first Type III TA system identified ($ToxIN_{Pa}$) was encoded by a cryptic plasmid (pECA1039) in *Pectobacterium atrosepticum*. $ToxIN_{Pa}$ was originally identified because ToxN shared 31% amino acid sequence identity with the AbiQ protein, which was involved in an abortive infection system in *Lactococcus lactis* (Emond et al., 1998; Fineran et al., 2009). Recent mutational analysis of the antitoxin and structural characterization of the toxin of the AbiQ system revealed that it is also a member of the Type III TA systems (Samson et al., 2013; Bélanger and Moineau, 2015). Since the discovery of $ToxN_{Pa}$, numerous other Type III TA systems have been identified bioinformatically in diverse bacterial genera and they can be chromosomally, plasmid, or phage-encoded (Blower et al., 2012). Type III TA systems have been classified into ToxIN, CptIN, and TenpIN families. The toxic proteins of these families are homologous to $ToxN_{Pa}$, while the antitoxin primary DNA sequences vary in length and number of tandem repeats. Interestingly, while $ToxIN_{Pa}$ exhibited a strong Abi phenotype, $ToxIN_{Bt}$ from *Bacillus thuringiensis* did not exhibit an Abi phenotype when challenged with over 100 different phages (Blower et al., 2012).

$ToxIN_{Pa}$ is composed of an RNA antitoxin, ToxI, which binds to and suppresses the toxic endoribonuclease protein, ToxN. Crystallographic evidence revealed that $ToxIN_{Pa}$ forms a triangular heterohexameric structure with 3 ToxN proteins in complex with 3 ToxI RNA pseudoknots (Blower et al., 2011). The complex is held in an inactive state under normal cellular conditions. However, during infection with specific phages, the complex is activated by an unknown mechanism, allowing the endoribonuclease to degrade host RNAs, leading to bacteriostasis and subsequent cell death. This phenomenon has features of a prokaryotic apoptosis that manifests itself in precocious death of the virally-infected bacterial host and, consequently, also inhibits progeny phage production. As this outcome restricts or terminates further phage invasion of the clonal bacterial population, the abortive infection process can be viewed as an altruistic suicide (Fineran et al., 2009).

Some phages aborted by $ToxIN_{Pa}$ have the capacity to evolve spontaneous resistant mutants at low frequency that can circumvent the Abi system. The $ToxIN_{Pa}$ resistant mutants of ΦTE, a *P. atrosepticum* phage, were found to overcome abortive infection by an RNA-based molecular mimicry of the ToxI antitoxin. The ΦTE escape phage had expanded a "pseudo-ToxI" region in the viral genome that was similar, but not identical, to the ToxI sequence. This expanded "pseudo-ToxI" region was expressed during phage infection and actively suppressed ToxN, thus allowing the ΦTE mutants to evade abortive infection (Blower et al., 2011). Recently, AbiQ-resistant mutations of four phages were sequenced revealing multiple loci involved in resistance, but how these genes conferred resistance to AbiQ remained elusive (Samson et al., 2013).

To date, ΦTE remains the only phage whose $ToxIN_{Pa}$-escape mechanism is understood. Thus, the primary aim of this study was to characterize additional resistance mutations of new $ToxIN_{Pa}$-sensitive phages and, by their study, perhaps add to the repertoire of known escape loci. Depending on the nature of the relevant mutation(s), the escape locus might provide insight into how phage infection leads to activation of the $ToxIN_{Pa}$ complex.

In this study, we isolated and characterized *Serratia* sp. ATCC 39006 (*S* 39006)-specific phages, ΦCHI14, ΦX20, and ΦCBH8. Comparison of the genomic sequences of spontaneous $ToxIN_{Pa}$-escape mutants and their wild type progenitors revealed three different routes of escape: (1) mutation of *asiA*, encoding a predicted phage transcriptional co-activator in ΦCHI14 and ΦCBH8; (2) mutation of an unknown gene (*orf84)* in a ΦCHI14 mutant; (3) deletion of a large region (6.5–10 kb) of the phage genome in most mutants of all three T4-family phages.

MATERIALS AND METHODS

Bacterial Strains, Bacteriophages, and Growth Conditions

Bacterial hosts and phages used in this study are listed in **Table 1**. All experiments were performed with *S* 39006. Bacteria were cultured at 30°C in Luria-Broth (LB) (10 g liter^{-1} tryptone, 5 g

TABLE 1 | Bacterial strains, plasmids, and bacteriophages used in this study.

Bacterial strain, plasmid, or phage	Relevant characteristics	References
STRAINS		
Serratia sp. ATCC 39006 LacA (wt)	Laboratory strain, referred to as wild type (wt) in text, Lac$^-$ derivative of *S* 39006, carbapenem+, prodigiosin+	
PLASMIDS*		
pTA46	$toxI_{Pa}$, $ToxN_{Pa}$	Fineran et al., 2009
pTA47	$toxI_{Pa}$, $ToxN_{Pa}$-frameshift (FS)	Fineran et al., 2009
pFR2	$tenpI_{Pl}$, $tenpN_{Pl}$	Blower et al., 2012
pFR8	$tenpI_{Pl}$, $tenpN_{Pl}$-FS	Blower et al., 2012
BACTERIOPHAGES		
ΦCHI14	Environmentally isolated phage	This study
ΦX20	Environmentally isolated phage	This study
ΦCBH8	Environmentally isolated phage	This study

**All plasmids containing Type III TA loci are in a pBR322 vector with ampicillin resistance for selection. The TA complexes are expressed from their native promoter.*

liter^{-1} yeast extract, 5 g liter^{-1} NaCl) or on LB agar (LBA) containing 1.5% w v^{-1} or 0.35% w v^{-1} agar to make LBA or top-LBA plates respectively. Bacterial growth was monitored by measuring optical density at 600 nm (OD_{600}) using a Thermo Scientific Helios Zeta spectrophotometer. Where required, media were supplemented with ampicillin at 100 μg mL^{-1}. Bacteriophages ΦCHI14, ΦX20, and ΦCBH8 were isolated from treated effluent collected from a sewage treatment plant in Cambridge, United Kingdom. The bacteriophages were selected from a library of *S* 39006-specific phages isolated using an enrichment procedure detailed previously (Evans et al., 2010). Phage lysates were generated as described previously (Petty et al., 2006). Spot tests were performed as described previously (Evans et al., 2010). Phages were stored at 4°C in phage buffer containing 10 mM $MgSO_4$, 10 mM Tris-HCl, 0.01% w v^{-1} gelatin and a few drops of chloroform. Efficiency of Plating (E.O.P.) was calculated after incubating serial dilutions of phage lysates overnight on bacterial lawns on LBA and dividing the titer of the phage on the test host by the titer of the phage on the control host (Kutter, 2009).

Isolation of Phage Escape Mutants

Plaques of rare spontaneous phage mutants were isolated on lawns of *S* 39006 cells expressing $ToxIN_{Pa}$ infected with wild type ΦCHI14, ΦX20, or ΦCBH8. Individual plaques were then purified at least twice on a lawn of *S* 39006 expressing $ToxIN_{Pa}$. Final lysates of escape mutant phages were then prepared from near-confluent lawns until a final titer of $>10^9$ plaque forming units (p.f.u.) mL^{-1} was obtained.

Electron Microscopy

Transmission Electron Micrograph (TEM) images of phages were taken at the Multi-Imaging Center, University of Cambridge using a Tecnai G2 series transmission electron microscope. Samples were prepared by adsorbing 10 μl of phage lysate ($>10^8$ p.f.u. mL^{-1}) onto a charged copper grid for 3 min. The grids were then washed with water twice before being stained with 2% phosphotungstic acid (PTA) neutralized with potassium hydroxide (KOH). The accelerating voltage was 120.0 kV and the direct magnification used to image phages was 25,000x.

Phage Genome Sequencing

Phage DNA was extracted using a standard phenol-chloroform protocol (Sambrook, 1989). In a phase-lock gel (PLG) tube (5′ Prime), 450 μL of high titer phage lysate was incubated with 4.5 μL of 1 mg/mL DNase I and 2.5 μL of 10 mg/mL RNase A and incubated at 37°C for 30 min. The mixture was then added to 11.5 μL of 20% SDS and 4.5 μL of 10 mg/mL Proteinase K and incubated for another 30 min. DNA was extracted by adding 500 μL of a Phenol:Chloroform:Isoamyl Alcohol 25:24:1 mix and centrifuged at 1,500 × g for 5 min. The supernatant was transferred to a new PLG tube and the previous step repeated. In a new PLG tube, the supernatant was supplemented with 500 μL of Chloroform:Isoamyl Alcohol 24:1 and centrifuged at 1,500 × g for 5 min. The aqueous phase at the top was then incubated with 45 μL sodium acetate (3 mol/L, pH 5.2) and 500 μL of 100% Isopropanol at room temperature for 15 min. The mixture was then subjected to centrifugation at 12,000 × g for 20 min, after which the pellet was washed at least twice with 70% ethanol and then re-suspended in dH_2O.

The genomes of wild type phages and selected escape mutants were sequenced using the Junior Roche 454 Genome Sequencer FLX pyrosequencer at the Department of Biochemistry, University of Cambridge or using the Illumina MiSeq, and HiSeq 2500 platforms at MicrobesNG. The resulting contigs were assembled using Genomic Sequencer *de novo* assembler (Roche) or SPAdes. The wild type sequences had coverage ranging from 57x to 100x of the full genome while the escape phage sequences had coverage ranging from 45x to 150x.

Genome Annotation and Bioinformatics

Phage genome open reading frames (ORFs) were defined using the gene prediction tools GeneMark.hmm (Lukashin and Borodovsky, 1998) and Glimmer (Delcher et al., 1999). Homologs of predicted proteins were identified using PSI-BLASTp searches or i-TASSER (Roy et al., 2010). The program tRNAScan-SE (Lowe and Eddy, 1997) was used to identify phage tRNA genes and ARAGORN (Laslett and Canback, 2004) was used to predict host tRNA genes. The genomic sequences of wild type and escape phages were compared using Geneious 6.1 (Biomatters Ltd) and Artemis (Rutherford et al., 2000). Protein alignments were conducted using the EMBOSS "ClustalW" program. Final alignment images were generated using the ESPript 2.2 program. All the above analysis were used at default settings.

DNA Manipulations

The *asiA* locus of additional ΦCHI14 and ΦCBH8 mutants was probed by PCR amplification using primers oBH3 (5′- CTGTGA CTTCGAGCTTAAATCTCC-3′) and oBH4 (5′- CGCTATATG TCAACAGGCCG-3′). Subsequent amplicons were subjected to Sanger sequencing.

Phage Burst Size

Phage burst size assays were performed as described previously (Petty et al., 2007). In brief, an overnight *S* 39006 culture was used to inoculate LB in a 250 mL conical flask and incubated at 30°C to $OD_{600} = 0.5$. Phage samples were then added at a multiplicity of infection (M.O.I.) of 0.001 and the culture incubated with shaking at 150 rpm at 30°C. Samples were taken at different time points and chloroform-treated before titrating to determine the number of p.f.u. The one-step growth curve describes phages per initial infection center, over time.

Phage Adsorption

An overnight culture of *S* 39006 was adjusted to $OD_{600} = 1$ with LB in a 250 mL conical flask and infected with phages at an M.O.I. of 0.001. The 10 mL infected culture was placed in a shaking water bath at 30°C with shaking at 150 rpm. One hundred microliters samples were taken at different time points and added to 900 μL of chilled LB. The samples were chloroform-treated immediately and then titrated. The final adsorption curve was plotted by calculating the percentage of free phages in the culture against

time. An LB-only sample was infected with phages as a negative control.

RESULTS

Three ToxIN$_{Pa}$-Sensitive *S* 39006 Phages Were Isolated

Since 2013, three phages that could be aborted by ToxIN$_{Pa}$ were isolated from water from treated sewage effluent in samples taken at least a year apart from each other. The phages were initially isolated after enrichment on the *S* 39006 host expressing ToxIN$_{Pa}$ with a frameshift mutation in the *toxN* gene. Abi sensitivity of these three phages, named ΦCHI14, ΦX20, and ΦCBH8 in chronological order, were examined initially by comparing titers from spot tests on *S* 39006 lawns expressing ToxIN$_{Pa}$ or with the frameshifted version of the ToxIN$_{Pa}$ locus as negative control. E.O.P. measurements showed that the three phages were strongly aborted by ToxIN$_{Pa}$: the E.O.P.s of ΦCHI14, ΦX20, and ΦCBH8 were 2.0×10^{-7}, 2.8×10^{-8}, and 8.0×10^{-6} respectively. Therefore, all three phages could produce "spontaneous escape" mutants that became insensitive to ToxIN$_{Pa}$ at low frequencies. Interestingly, all three phages were also aborted by another Type III TA system (TenpIN$_{Pl}$) (Blower et al., 2012) but without producing any detectable spontaneous escape mutants (E.O.P. $<10^{-9}$). None of the phages showed sensitivity to other Type III TA systems tested.

ΦCHI14, ΦX20, and ΦCBH8 Are T4-Like Phages of the Myoviridae Family

TEM images revealed that all three phages had isometric, icosahedral heads, contractile tails and tail fibers (**Figure 1**). This classified them in the Caudovirales order and Myoviridae family (Ackermann, 2009). Whole genome sequencing results revealed that ΦCHI14, ΦX20, and ΦCBH8 were very similar to each other at the DNA sequence level. ΦCHI14 and ΦCBH8 both have a size of 171,151 bp and contain 275 predicted open reading frames (ORFs). ΦCHI14 and ΦCBH8 are almost identical genomically, except for 19 point mutations, 11 of which are in ORFs and cause non-synonymous mutations. tRNAscan-SE identified 16 tRNA genes encoded by ΦCHI14 and ΦCBH8. Similarly ΦX20 has a genome of 172,450 bp with 279 ORFs plus 17 predicted tRNA genes and it shared 93.8% homology with the genomes of ΦCHI14 and ΦCBH8. Interestingly all three phages encode almost twice the number of tRNAs as other related phages, such as T4 and CC31. Further, all three phages also have a GC content of ∼38%, much lower than the host GC content of 49.24%. The genomes of all three phages were deposited in GenBank with the following accession numbers: ΦCHI14 (MF036690), ΦCBH8 (MF036691), and ΦX20 (MF036692).

ΦCHI14 was used as a representative of the new isolates to compare with related phages (**Figure 2**): ΦCHI14 shares 55.0, 57.5, 55.3, 54.3, 57.4, and 60.1% DNA sequence identity with phages T4, CC31, ΦR1-RT, ΦS16, PG7, and PEi20 respectively—all of the above being T4-like phages.

ToxIN$_{Pa}$-Sensitive *S* 39006 Phages Harbor at Least Three Distinct ToxIN$_{Pa}$ Escape Loci

Although ToxIN$_{Pa}$ aborted T4-like *S 39006* phages efficiently, rare phage plaques appeared at low frequencies. We presumed that these rare plaques arose due to viral mutations enabling phages to "escape" or circumvent the effects of ToxIN$_{Pa}$. Therefore, we expected that definition of the corresponding mutations would enable identification of the phage genes encoding products responsible for "activation" of the ToxIN$_{Pa}$ system.

In total, 24 "escape" mutants of these *S* 39006 phages were subject to whole genome sequencing; 5 of ΦCHI14 mutants, 11 of ΦX20 mutants and 8 of ΦCBH8 mutants. Mutations were

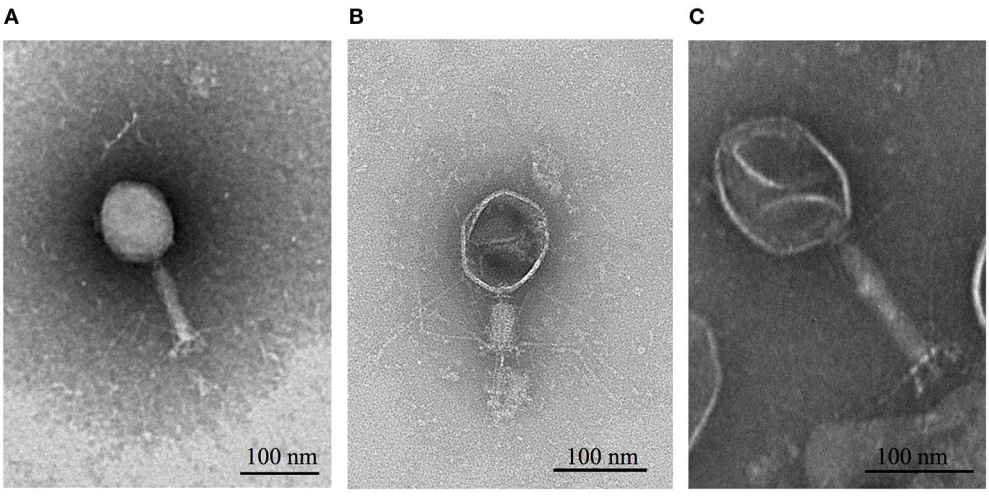


FIGURE 1 | TEM images of **(A)** ΦCHI14, **(B)** ΦX20, and **(C)** ΦCBH8. ΦCHI14 and ΦCBH8 have extended tails while ΦX20 captured in the images shows a contracted tail.

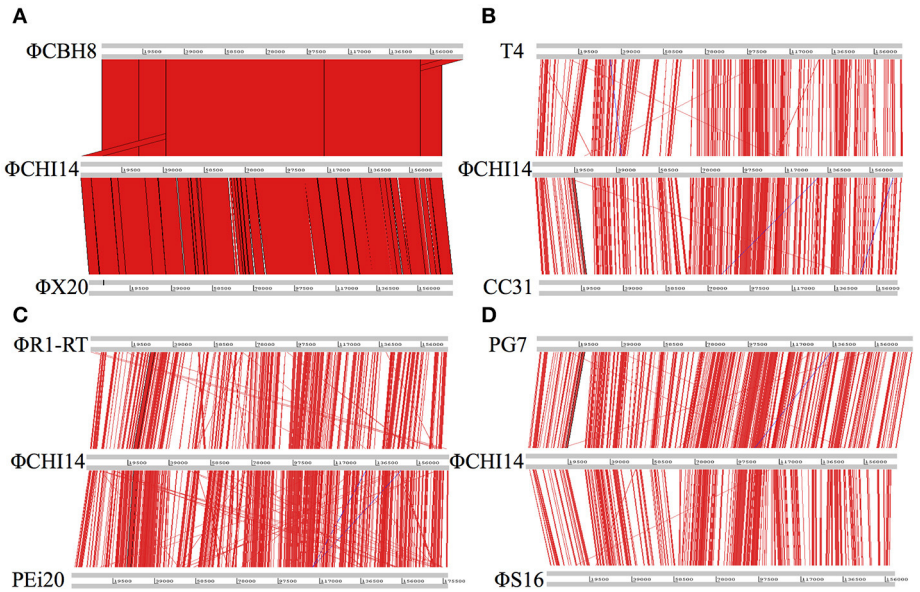


FIGURE 2 | Whole genome alignment of wild type ΦCHI14 with other T4-like phages. DNA homologies are displayed in red. The degree of similarity is proportional to the intensity of red. Similar linear organization of the genome is observed between the aligned phages: **(A)** Comparison with ΦCBH8 and ΦX20. **(B)** Comparison with T4 and CC31. **(C)** Comparison with phage ΦR1-RT and PEi20. **(D)** Comparison with PG7 and ΦS16.

mapped to the corresponding wild type genomes to identify the putative ToxIN$_{Pa}$ resistance loci. Interestingly, the mutations did not all map to single locations in the corresponding genomes. Instead, three distinct mutational types were identified in the genomes of the 24 escape phages (**Table 2**): 21 mutants harbored a large deletion (size ranging from 6.5 to 10 kb) in their genomes; 2 mutants carried a nonsense mutation or deletion in the *asiA* gene; and 1 mutant had a missense mutation in an *orf* encoding a hypothetical protein, ORF84.

Asia is Involved in the Activation of ToxIN$_{Pa}$

Two sequenced escape mutants of ΦCHI14 (ΦCHI14b and ΦCHI14f) had mutations in the *asiA* gene (**Table 2**). After defining these mutant *asiA* variants, the *asiA* locus in additional ΦCHI14 and ΦCBH8 "escape" mutants were probed by PCR and Sanger sequencing and 6 further independent *asiA* mutants were identified. The *asiA* locus encodes the protein AsiA, the homolog of which in ΦT4 is involved in σ^{70}-appropriation (Hinton et al., 2005). Most of the mutations affect the C-terminal domain (CTD) of AsiA, including: (1) truncation of the CTD via a point mutation leading to a premature stop codon; (2) extension of the CTD via insertion or deletion of nucleotides leading to frameshift mutations that eliminated the natural stop codon. Only one mutant had a V13G mutation in the N-terminal domain of the protein (**Figure 3**).

ORF84 May Be Involved in the Activation of ToxIN$_{Pa}$

One mutant (ΦCHI14e) had a single nucleotide substitution in the gene, *orf84*. The A to C substitution caused an E66D change in the encoded ORF84 protein. Homologs of ORF84 have been found in 11 other phages, the most similar homolog belonging to enterobacterial phage, CC31. However, no functional information exists currently for any of the homologs from both sequenced-based predictions and structure-based predictions.

The "Large Deletion" is the Most Prevalent Mutational Route through Which These *S* 39006 Phages Escape ToxIN$_{Pa}$

The majority (21 out of 24) of the sequenced "escape" mutants of *S* 39006 phages became insensitive to ToxIN$_{Pa}$ by a similar type of mutation—through a large deletion of a specific viral genome locus (**Table 2**). This was the only common mutational route through which all three *S* 39006 phages could escape ToxIN$_{Pa}$. In particular, all 11 of the ΦX20 mutants isolated arose through this "large deletion" mutation. The corresponding deletions overlapped across a common core and varied from 6,521 bp (ΦCBH8o) to 10,094 bp (ΦCHI14c). The largest deleted region contains 19 ORFs and 13 tRNA genes, whereas the smallest contains 13 ORFs and 10 tRNA genes (**Figure 4**). Although the precise 5′ and 3′ borders of most deletion mutations were variable, closer inspection revealed the presence of direct repeats flanking the deleted region in every mutant (**Table 3**). The repeat length and sequence was unique in each mutant and most of the repeats appeared numerous times in the wild type genome, some are also represented within the deleted region. The presence of the direct repeats may suggest slipped mispairing during replication, or inter/intra-molecular misalignment during recombination, as possible mechanisms driving the deletion mutations (Singer and Westlye, 1988; Pierce and Masker, 1989).

TABLE 2 | Summary of mutations in ΦCHI14, ΦX20, and ΦCBH8 ToxIN$_{Pa}$ escape mutants.

Escape phage	Mutation type	Gene(s) affected	Effect of mutation
ΦCHI14a	large deletion (7,647 bp)	14 ORFs 12 tRNA genes	elimination of affected genes
ΦCHI14b	nonsense mutation	*asiA*	E71 ➔ stop codon
ΦCHI14c	large deletion (10,094 bp)	19 ORFs 13 tRNA genes	elimination of affected genes
ΦCHI14e	missense mutation	*orf84*	E66D
ΦCHI14f	deletion (10 bp)	*asiA*	extends protein by 17 residues
ΦCBH8f	large deletion (10,040 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦCBH8l	large deletion (7,802 bp)	15 ORFs 11 tRNA genes	elimination of affected genes
ΦCBH8m	large deletion (8,575 bp)	17 ORFs 6 tRNA genes	elimination of affected genes
ΦCBH8o	large deletion (6,521 bp)	14 ORFs 10 tRNA genes	elimination of affected genes
ΦCBH8p	large deletion (8,368 bp)	16 ORFs 10 tRNA genes	elimination of affected genes
ΦCBH8t	large deletion (7,328 bp)	15 ORFs 9 tRNA genes	elimination of affected genes
ΦCBH8u	large deletion (8,158 bp)	17 ORFs 4 tRNA genes	elimination of affected genes
ΦCBH8x	large deletion (7,731 bp)	14 ORFs 10 tRNA genes	elimination of affected genes
ΦX20b	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20d	large deletion (9,479 bp)	19 ORFs 8 tRNA genes	elimination of affected genes
ΦX20f	large deletion (9,473 bp)	19 ORFs 8 tRNA genes	elimination of affected genes
ΦX20g	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20h	large deletion (9533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20j	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20k	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20l	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20m	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20n	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes
ΦX20o	large deletion (9,533 bp)	19 ORFs 11 tRNA genes	elimination of affected genes

Deleted Regions Contain Mostly Unknown ORFs

The ability of "large deletion" mutants to escape ToxIN$_{Pa}$ infers that the 6.5–10 kb deleted region contains genetic elements directly or indirectly responsible for activating ToxIN$_{Pa}$. In our limited pool of spontaneous "large deletion" mutants, 6.5 kb was the most compact deletion and therefore detailed inspection of the ORFs and tRNAs in this region was undertaken. Every individual predicted ORF was investigated by searching for DNA homologies using nucleotide BLAST; amino acid sequence homology using protein BLAST; and finally i-TASSER for predicted structural homologs.

Of the 13 predicted ORFs within the smallest deletion region found in ΦCBH8o, 6 encode hypothetical proteins with no sequence, or structural, homologs, while 7 of the ORFs encode homologs in other T4-like phages. However, no functional information is available about any of these homologs—except for one hypothetical protein encoded by the 3′ end of the deletion region. This hypothetical protein, (ORF145), shows some similarity to the membrane anchor domain of an agglutinating adhesin (YadA) in both protein BLAST and i-TASSER structural predictions.

The absence of any functional information for individual ORFs in the "large deletion" locus prompted an analysis of whether the region is present in other related phages or is unique to the T4-family *S* 39006 phages reported in this study. Therefore, the 6.5 kb smallest deletion region from ΦCBH8o was aligned with the genomes of the related phages that show high homology with the entire genomes of these environmental T4-like *S* 39006 phages. ACT alignment showed only very limited identity of the 6.5 kb region with phages T4, CC31, ΦR1-RT, ΦS16, PG7, and PEi20 (**Figure 5**).

The presence of multiple tRNA genes within the deleted regions may also suggest that these tRNAs play a role in susceptibility to ToxIN$_{Pa}$. Among all the "large deletion" mutants, ΦCBH8u showed deletion of the smallest number of tRNAs: Gly with the TTC anticodon, Met with the CAT anticodon, Arg with the TCT anticodon and Leu with the TAA anticodon. tRNAs genes with the same anticodons for the above mentioned amino acids appear 3 times and 6 times for the first two and only once for the last two in the host genome. The presence of these tRNA genes in the host genome suggests it may be unlikely that they are the direct activators of ToxIN$_{Pa}$. However, the low frequency of Arg (TCT) and Leu (TAA) in the host genome suggests the possibility that their deletion from the phage genome might lead to inefficient translation of some viral transcripts encoding products that activate ToxIN$_{Pa}$.

Loss of the Large Deletion Locus Affects the Fitness of Mutant Phages

Given the extent of the viral genome deletions and absence of any functional information on most of the deleted gene products, we decided to investigate the impact of the deletions on phage fitness, represented by adsorption efficiency and burst size. The burst size of wild type ΦCBH8 and ΦCBH8o infecting exponential phase *S* 39006 with an M.O.I. of 0.001 was measured by one-step growth ($n = 5$). Both wild type ΦCBH8 and ΦCBH8o showed a latent period of 25 min and a rise period of about 35 min. The average burst size was

```
CBH8wt   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14wt  1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14b   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14f   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14q   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14s   1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CHI14w   1 MSKIEIVREIVTGASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8a    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8l    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60
CBH8r    1 MSKIEIVREIVTVASILIKTSCEDILEKRENFIAFLNELGLRNEHGRELNLANFKKMIDG 60

CBH8wt  61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA---------------------------- 89
CHI14wt 61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA---------------------------- 89
CHI14b  61 LNDDERSSLV------------------------------------------------ 70
CHI14f  61 LNDDERSSLVEEFNEGFEDIWLCITRKGDRSLLILLFLKYGRPVDI------------ 106
CHI14q  61 LNDDERSSLVEEFNEGFEDISIVIWLCITRKGDRSLLILLFLKYGRPVDI--------- 110
CHI14s  61 LNDDERSSLVEEFNEGFEDIYIVIWLCITRKGDRSLLILLFLKYGRPVDI--------- 110
CHI14w  61 LNDDERSSLVEEFNEGFEDIYRHLAMHNA---------------------------- 89
CBH8a   61 LNDDERSSLV------------------------------------------------ 70
CBH8l   61 LNDDERSSLVE----------------------------------------------- 71
CBH8r   61 LNDDERSSLVE----------------------------------------------- 71
```

FIGURE 3 | Alignment of the primary sequence of AsiA encoded in ΦCHI14 and ΦCBH8 wild type and escape phages. Blue highlight indicates identical sequences while yellow highlight shows variations in the primary sequence of AsiA in escape mutants.

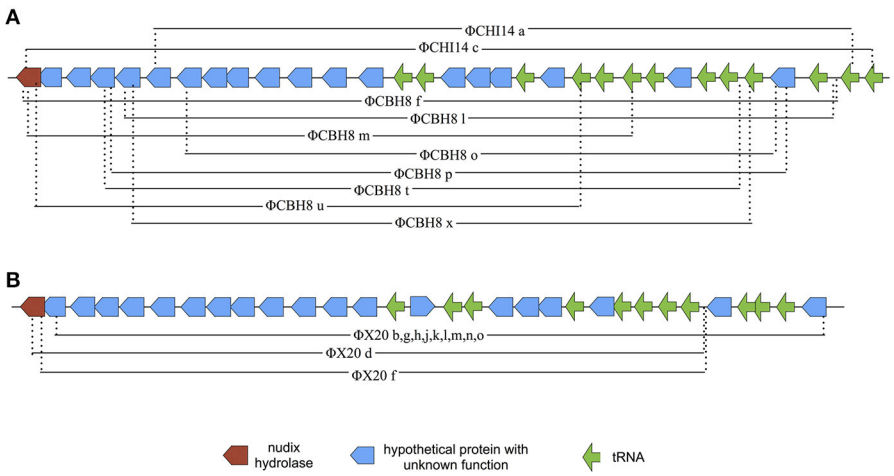


FIGURE 4 | Diagram of the deleted regions in ΦCHI14, ΦCBH8, and ΦX20 mutants. **(A)** Mapping of "large deletion" regions onto ΦCHI14 and ΦCBH8 genomes shows the start and end of each deletion. The same presentations of the deleted regions are used for ΦCHI14 and ΦCBH8 due to their high similarity. **(B)** Mapping of "large deletion" regions onto the ΦX20 genome shows the start and end of each deletion.

approximately 22 phage particles per initial infection center for wild type ΦCBH8 and 28 for ΦCBH8o (**Figure 6A**). However, the adsorption efficiency of ΦCBH8o was lower than that of wild type ΦCBH8 (**Figure 6A**), suggesting a fitness defect. To investigate if the difference in adsorption was phage-dependent or host-dependent, we carried out a second adsorption assay ($n = 3$) using stationary phase *S* 39006 instead of exponential phase host but retaining the same M.O.I. As shown in **Figure 6B**, no obvious difference in binding to host cells was observed—both phages achieving >95% adsorption in 20 min. These results may suggest that the "large deletion" mutation causes decrease in adsorption efficiency only when infecting exponential phase host cultures, but causes no significant difference in burst size. However, due to the variable nature of the fitness assays, further experiments should be carried out in the future to verify the differences in adsorption seen in ΦCBH8 and the ΦCBH8o mutant, especially in the exponential phase of bacterial growth.

DISCUSSION

In this study we isolated three highly homologous T4-family environmental phages of *S* 39006 that are sensitive to $ToxIN_{Pa}$-mediated Abi. Rare spontaneous mutants of these phages were able to circumvent $ToxIN_{Pa}$ via multiple escape routes. The independent isolation of ΦCHI14, ΦX20, and ΦCBH8

TABLE 3 | Sequences of direct repeat flanking the "large deletion" region in mutants are shown.

Mutant name	Repeat sequence	Frequency of repeat in wild type genome	Frequency of repeat within deleted region
ΦCHI14a	TCAGCCA	40	3
ΦCHI14c	GGATTA	>100	4
ΦCBH8f	GAACTGC	22	3
ΦCBH8l	TTGAGTAG	7	0
ΦCBH8m	GTCCCTG	11	0
ΦCBH8o	CCGAAGC	15	1
ΦCBH8p	GTTCAC	94	4
ΦCBH8t	AGCCATCC	5	0
ΦCBH8u	GGAAGCC	22	1
ΦCBH8x	ATCTG	>100	11
ΦX20b, g, h, j, k, l, m, n, o	AACTGCTACA	2	0
ΦX20d	GGGAAAC	6	0
ΦX20f	CTTCGCC	21	1

The repeat length varies from 5 to 10 nucleotides and the sequences are different in each mutant of unique deletion size. Most of the repeats appear numerous times in the wild type genome and some of them appear also within the deleted region.

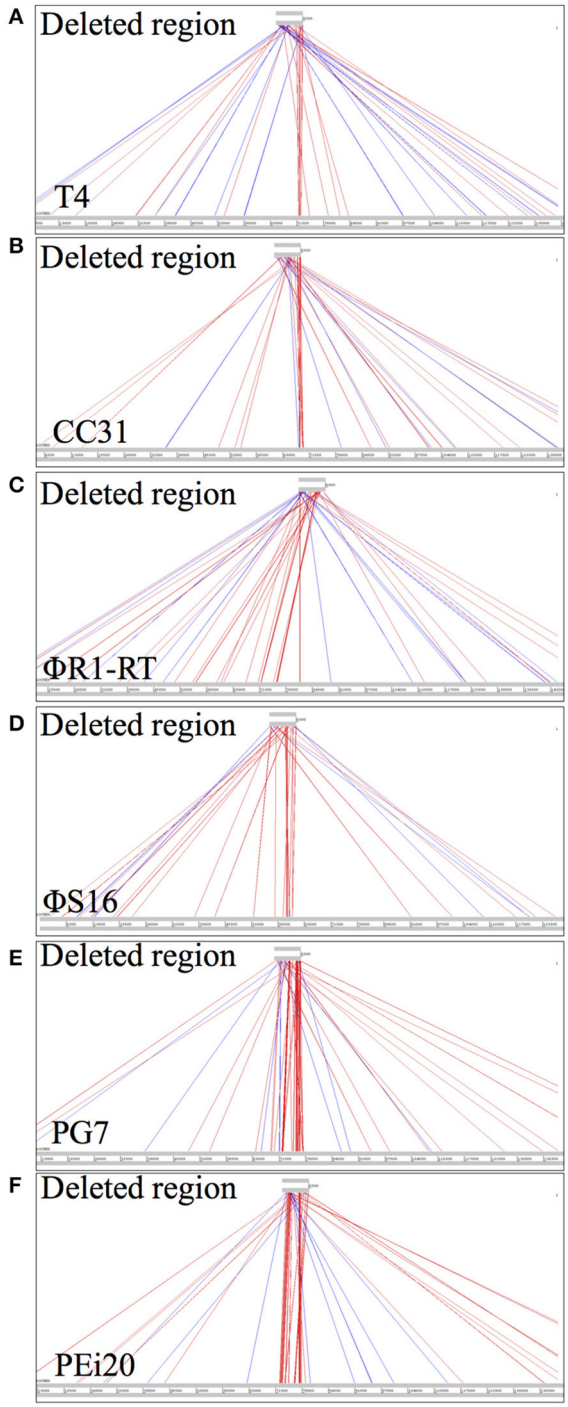


FIGURE 5 | DNA sequence comparisons of the smallest deleted region in ΦCBH8o (6.5 kb) with other T4-like phages, using ACT. **(A)** Comparison with phage T4. **(B)** Comparison with CC31. **(C)** Comparison with ΦR1-RT. **(D)** Comparison with ΦS16. **(E)** Comparison with PG7. **(F)** Comparison with PEi20.

is interesting since pairwise alignment of their genomes with other T4-like phages showed that they are mosaics of one another—whole-genome alignments consist of varying-length stretches of high homology interspersed with stretches of no homology (Nolan et al., 2006; Petrov et al., 2006). This kind of similarity and diversity is typical among T4-like phages (Petrov et al., 2010) yet the relatively low overall sequence similarities indicate that ΦCHI14, ΦX20, and ΦCBH8 may represent a new type of T4-like genome configuration. Furthermore, the isolation of only 3 ToxIN$_{Pa}$-sensitive, and very similar phages (despite multiple environmental enrichments yielding more than 300 phages over a 3-year span) suggests a relatively rare incidence of ToxIN$_{Pa}$-sensitive viruses among environmental phages of *S* 39006. However, this rarity may not be surprising because wild type *S* 39006 does not carry ToxIN$_{Pa}$ naturally. Furthermore, it is possible that these new phages arose in the environment by propagation on alternative native hosts taxonomically unrelated to *S* 39006.

We showed recently that a single phage gene product from phage ΦM1 was responsible for activating the ToxIN$_{Pa}$ system in the natural host, *P. atrosepticum* (Blower et al., 2017). A phage product might interact with ToxI and degrade it or sequester it away from ToxN to liberate the toxin. Alternatively, a phage product could interact with ToxN, reducing the affinity between ToxN and ToxI. Interactions of the phage product(s) with both ToxI and ToxN are also formally possible. Based on our recent observations with the evolution of ΦM1 to ToxIN$_{Pa}$ resistance in *P. atrosepticum*, we expected to find mutations in a single locus in "escape" mutants of ΦCHI14, ΦX20, and ΦCBH8. However, our characterization of the latter "escape" mutants has now suggested a more complicated landscape for potential activation mechanisms operating on ToxIN$_{Pa}$.

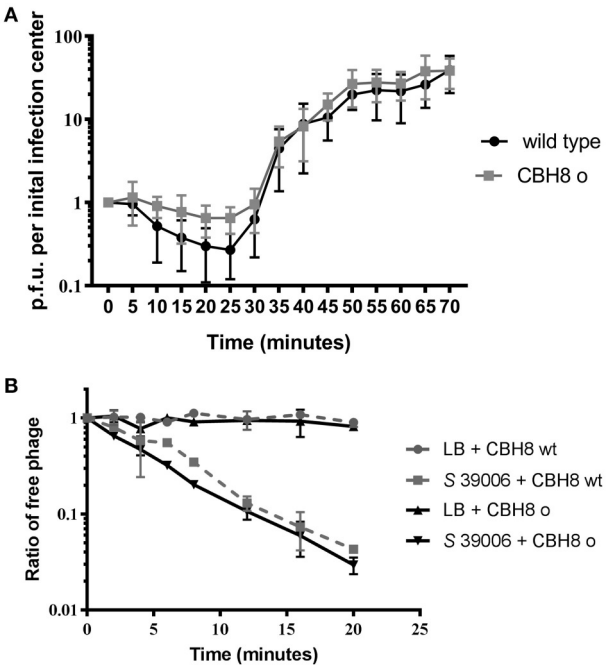


FIGURE 6 | Fitness comparison of ΦCBH8wt and ΦCBH8o. **(A)** One-step growth curve of ΦCBH8wt and ΦCBH8o. **(B)** Adsorption of ΦCBH8wt and ΦCBH8o to wild type *S* 39006. ΦCBH8wt and ΦCBH8o with LB-only was used as a negative control for adsorption.

One "escape" locus found in both ΦCHI14 and ΦCBH8 mutants is *asiA*, which encodes AsiA, a highly-conserved protein in T4-like phages (**Figure 7**). In T4, AsiA is an anti-sigma factor that represses host transcription through a process known as σ-appropriation, and, together with MotA, co-activates transcription of phage middle genes (Hinton et al., 2005). AsiA monomers bind tightly to regions 4.1 and 4.2 of σ^{70} and abolish the sigma factor's ability to bind to the −35 promoter sequence of host genes (Hinton, 2010). As a result, AsiA inhibits transcription of bacterial genes with a −10/−35 promoter. In addition, AsiA-bound σ^{70} adopts an altered conformation that allows the T4 transcriptional activator, MotA, to bind σ^{70} and the MotA box present in the promoter region of T4 middle-expressed genes. Thus, the AsiA-σ^{70}-MotA complex disrupts the ability of RNA-polymerase (RNAP) to recognize and transcribe host genes but reconfigures the enzyme to transcribe phage middle-expressed genes (Ouhammouch et al., 1995; Lambert et al., 2004). Most of the 6 escape mutants of ΦCHI14 and ΦCBH8 produce an AsiA with either an extended or truncated CTD, while only one mutant has a V13G substitution in the NTD. In AsiA the role of the CTD is controversial and not fully understood, with some studies showing that mutations in the CTD of AsiA affected its function in σ-appropriation and MotA binding (Yuan and Hochschild, 2009; Yuan et al., 2009) while other studies showed no compromise in σ-appropriation of AsiA with a mutated CTD (Pal et al., 2003). As for the V13G mutation in the NTD, the mutated amino acid is homologous to V14 in T4, which was defined as an essential residue for AsiA-AsiA homodimer and AsiA-σ^{70} interactions (Gilmore et al., 2010). Future research will investigate whether AsiA plays a direct role in $ToxIN_{Pa}$ activation as a phage product or as a part of the AsiA-RNAP-MotA complex, or whether AsiA indirectly activates $ToxIN_{Pa}$ by disturbing ToxI:ToxN stoichiometry or phage middle gene transcription.

The most frequently found escape locus among the *S* 39006 phage mutants was the deletion of a substantial region of the phage genomes. The size of the deletion represented 3.8–5.8% of the total genome size. These deletant mutants remained viable on *S* 39006 indicating that the deleted region is not essential for phage replication, at least under the laboratory conditions used. How the deletion mutations confer resistance is not immediately obvious, given the number of genes affected. As no missense mutations were isolated that mapped to this locus, we presume that two or more genes have to be mutated in this locus simultaneously for phages to escape, and presumably, these genes might be located toward the two ends of the deletion, hence the large deletion size. An interesting hypothesis here would be that the entire deletion region might contain genes that wild type phages acquired by recombinogenic lateral gene transfer (Petrov et al., 2010) in the natural environment (perhaps during mixed viral infections) and these genes, in concert, could be responsible for activating $ToxIN_{Pa}$ in bacterial hosts. The possibility that the deleted region was acquired through lateral gene transfer is supported by two pieces of evidence: (1) attempts to align the large deletion locus with several other T4-like phages (T4, CC31, ΦR1-RT, ΦS16, PG7, and PEi20) showed very little homology (**Figure 5**); and (2) most of the 5' borders of the large deletion mutations lie within tRNA genes (**Figure 4**), whose conserved sequences are readily recombinogenic. It could be argued that this has echoes of the descriptions of Pathogenicity Islands in bacteria (Karaolis et al., 1999; Schmidt and Hensel, 2004) and therefore could even suggest that the large deletion region is a potentially mobile viral genome unit.

The only clue to possible functions of ORFs in the deleted regions comes from ORF145, which is a YadA homolog. YadA was first found in *Yersinia* species and serves as a virulence factor that mediates *Yersinia* adherence to epithelial tissue (Hoiczyk et al., 2000). YadA cognate genes have also been found in other phages but it is unknown if they might be transferred to their bacterial hosts to enhance virulence (Moreno Switt et al., 2013). Future work will involve investigating the impact of ORF145 in allowing phage "escape." Finally, given the important physiological roles for phage-encoded tRNAs in viral morphogenesis, future research into the role that the tRNAs encoded within the deleted regions may play in phage "escape" requires investigation.

In summary, this study has discovered a distinct group of T4-family *S* 39006 phages that could activate the $ToxIN_{Pa}$-mediated abortive infection system. The data on the "escape" mutants of these environmental phages suggest that either $ToxIN_{Pa}$ can be activated by more than one phage product, or that production of the $ToxIN_{Pa}$- activating product involves multiple biological processes and that defects in several could enable phages to circumvent Abi. Future investigations will

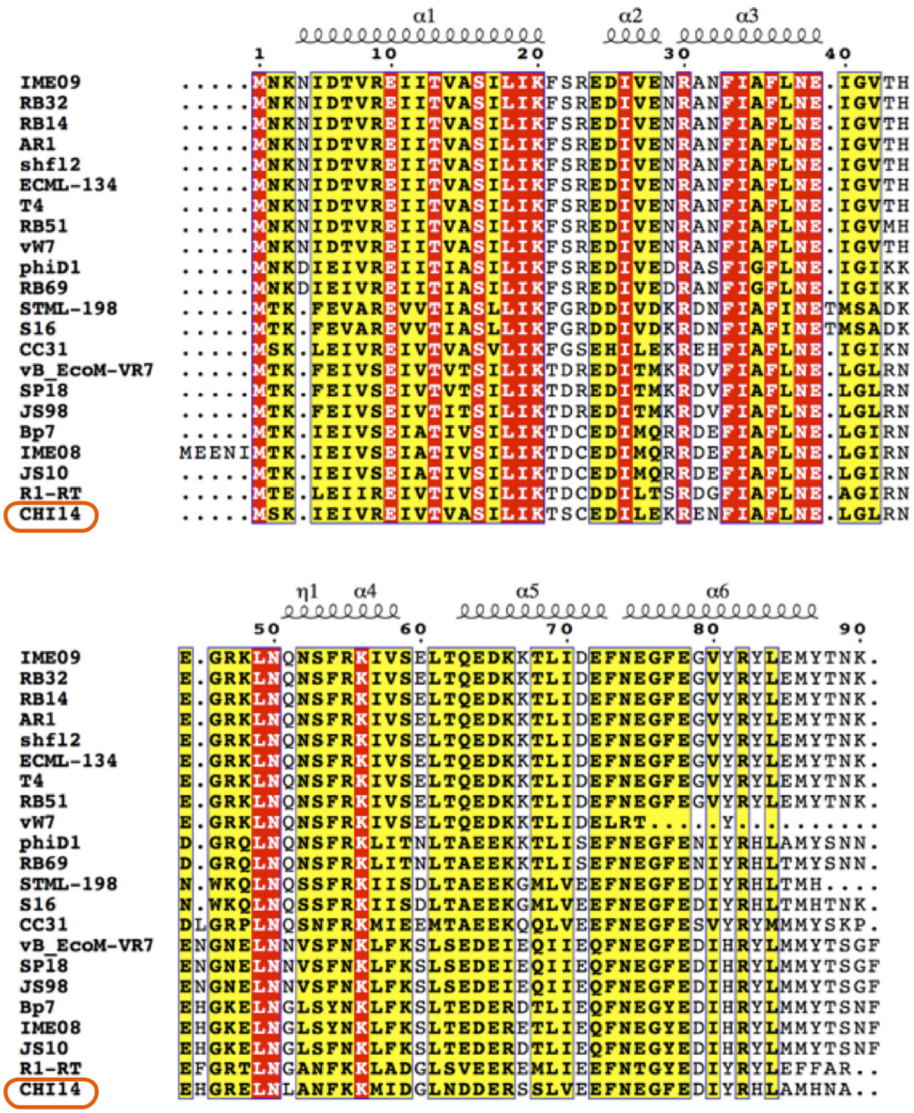


FIGURE 7 | Alignment of AsiA from ΦCHI14 and related phages. The N-terminal region of AsiA is highly conserved among phages while the C-terminal region is more varied. The protein secondary structure prediction is overlaid on the alignment.

involve further characterization of the products from the escape loci of these phages, and experiments to try to dissect their physiological roles during the process of abortive infection—and its circumvention.

AUTHOR CONTRIBUTIONS

GS, BC, CA, and XF conceived and designed the experiments. BC, CA, and XF performed the experiments. BC and CA prepared figures and graphs. BC and GS wrote the manuscript. All the authors read and approved the final manuscript.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC), UK. BC was funded by the Cambridge Commonwealth, European and International Trust. CA was funded by The Gates Cambridge Trust.

ACKNOWLEDGMENTS

We thank Dr. N. Goeders, Dr. R. E. Monson, C. Lee and R. Chai for helpful discussion of some experiments.

REFERENCES

Ackermann, H. W. (2009). Phage classification and characterization. *Methods Mol. Biol.* 501, 127–140. doi: 10.1007/978-1-60327-164-6_13

Bélanger, M., and Moineau, S. (2015). Mutational analysis of the antitoxin in the Lactococcal type III toxin-antitoxin system AbiQ. *Appl. Environ. Microbiol.* 81, 3848–3855. doi: 10.1128/AEM.00572-15

Blower, T. R., Chai, R., Przybilski, R., Chindhy, S., Fang, X., Kidman, S. E., et al. (2017). Evolution of *Pectobacterium* bacteriophage ΦM1 to escape two bifunctional type III toxin-antitoxin and abortive infection systems through mutations in a single viral gene. *Appl. Environ. Microbiol.* 83:e03229-16. doi: 10.1128/AEM.03229-16

Blower, T. R., Pei, X. Y., Short, F. L., Fineran, P. C., Humphreys, D. P., Luisi, B. F., et al. (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat. Struct. Mol. Biol.* 18, 185–190. doi: 10.1038/nsmb.1981

Blower, T. R., Short, F. L., Rao, F., Mizuguchi, K., Pei, X. Y., Fineran, P. C., et al. (2012). Identification and classification of bacterial type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. *Nucleic Acids Res.* 40, 6158–6173. doi: 10.1093/nar/gks231

Chopin, M. C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in *lactococci*: variations on a theme. *Curr. Opin. Microbiol.* 8, 473–479. doi: 10.1016/j.mib.2005.06.006

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641.

Emond, E., Dion, E., Walker, S. A., Vedamuthu, E. R., Kondo, J. K., and Moineau, S. (1998). AbiQ, an abortive infection mechanism from *Lactococcus lactis*. *Appl. Environ. Microbiol.* 64, 4748–4756.

Evans, T. J., Crow, M. A., Williamson, N. R., Orme, W., Thomson, N. R., Komitopoulou, E., et al. (2010). Characterization of a broad-host-range flagellum-dependent phage that mediates high-efficiency generalized transduction in, and between, *Serratia* and *Pantoea*. *Microbiology* 156(Pt 1), 240–247. doi: 10.1099/mic.0.032797-0

Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., and Salmond, G. P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.* 106, 894–899. doi: 10.1073/pnas.0808832106

Gerdes, K., Christensen, S. K., and Løbner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* 3, 371–382. doi: 10.1038/nrmicro1147

Gilmore, J. M., Bieber Urbauer, R. J., Minakhin, L., Akoyev, V., Zolkiewski, M., Severinov, K., et al. (2010). Determinants of affinity and activity of the anti-sigma factor AsiA. *Biochemistry* 49, 6143–6154. doi: 10.1021/bi1002635

Hinton, D. M. (2010). Transcriptional control in the prereplicative phase of T4 development. *Virol. J.* 7:289. doi: 10.1186/1743-422X-7-289

Hinton, D. M., Pande, S., Wais, N., Johnson, X. B., Vuthoori, M., Makela, A., et al. (2005). Transcriptional takeover by sigma appropriation: remodelling of the sigma70 subunit of *Escherichia coli* RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. *Microbiology* 151(Pt 6), 1729–1740. doi: 10.1099/mic.0.27972-0

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., and Heesemann, J. (2000). Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J.* 19, 5989–5999. doi: 10.1093/emboj/19.22.5989

Karaolis, D. K., Somara, S., Maneval, D. R., Johnson, J. A., and Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399, 375–379. doi: 10.1038/20715

Kutter, E. (2009). *Phage Host Range and Efficiency of Plating*. New York, NY: Humana Press.

Lambert, L. J., Wei, Y., Schirf, V., Demeler, B., and Werner, M. H. (2004). T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J.* 23, 2952–2962. doi: 10.1038/sj.emboj.7600312

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.

Lukashin, A. V., and Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* 26, 1107–1115.

Moreno Switt, A. I., Orsi, R. H., den Bakker, H. C., Vongkamjan, K., Altier, C., and Wiedmann, M. (2013). Genomic characterization provides new insight into *Salmonella* phage diversity. *BMC Genomics* 14:481. doi: 10.1186/1471-2164-14-481

Nolan, J. M., Petrov, V., Bertrand, C., Krisch, H. M., and Karam, J. D. (2006). Genetic diversity among five T4-like bacteriophages. *Virol. J.* 3:30. doi: 10.1186/1743-422X-3-30

Ouhammouch, M., Adelman, K., Harvey, S. R., Orsini, G., and Brody, E. N. (1995). Bacteriophage T4 MotA and AsiA proteins suffice to direct *Escherichia coli* RNA polymerase to initiate transcription at T4 middle promoters. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1451–1455.

Pal, D., Vuthoori, M., Pande, S., Wheeler, D., and Hinton, D. M. (2003). Analysis of regions within the bacteriophage T4 AsiA protein involved in its binding to the sigma70 subunit of *E. coli* RNA polymerase and its role as a transcriptional inhibitor and co-activator. *J. Mol. Biol.* 325, 827–841. doi: 10.1016/S0022-2836(02)01307-4

Petrov, V. M., Nolan, J. M., Bertrand, C., Levy, D., Desplats, C., Krisch, H. M., et al. (2006). Plasticity of the gene functions for DNA replication in the T4-like phages. *J. Mol. Biol.* 361, 46–68. doi: 10.1016/j.jmb.2006.05.071

Petrov, V. M., Ratnayaka, S., Nolan, J. M., Miller, E. S., and Karam, J. D. (2010). Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virol. J.* 7:292. doi: 10.1186/1743-422X-7-292

Petty, N. K., Foulds, I. J., Pradel, E., Ewbank, J. J., and Salmond, G. P. (2006). A generalized transducing phage (phiIF3) for the genomically sequenced *Serratia marcescens* strain Db11: a tool for functional genomics of an opportunistic human pathogen. *Microbiology* 152(Pt 6), 1701–1708. doi: 10.1099/mic.0.28712-0

Petty, N. K., Toribio, A. L., Goulding, D., Foulds, I., Thomson, N., Dougan, G., et al. (2007). A generalized transducing phage for the murine pathogen *Citrobacter rodentium*. *Microbiology* 153(Pt 9), 2984–2988. doi: 10.1099/mic.0.2007/008888-0

Pierce, J. C., and Masker, W. (1989). Genetic deletions between directly repeated sequences in bacteriophage T7. *Mol. Gen. Genet.* 217, 215–222.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. doi: 10.1038/nprot.2010.5

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944

Sambrook (1989). *Molecular Cloning; A Laboratory Manual*. New York, NY: Cold Spring Harbour Laboratory Press.

Samson, J. E., Bélanger, M., and Moineau, S. (2013). Effect of the abortive infection mechanism and type III toxin/antitoxin system AbiQ on the lytic cycle of *Lactococcus lactis* phages. *J. Bacteriol.* 195, 3947–3956. doi: 10.1128/JB.00296-13

Schmidt, H., and Hensel, M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* 17, 14–56. doi: 10.1128/CMR.17.1.14-56.2004

Singer, B. S., and Westlye, J. (1988). Deletion formation in bacteriophage T4. *J. Mol. Biol.* 202, 233–243.

Yuan, A. H., and Hochschild, A. (2009). Direct activator/co-activator interaction is essential for bacteriophage T4 middle gene expression. *Mol. Microbiol.* 74, 1018–1030. doi: 10.1111/j.1365-2958.2009.06916.x

Yuan, A. H., Nickels, B. E., and Hochschild, A. (2009). The bacteriophage T4 AsiA protein contacts the β-flap domain of RNA polymerase. *Proc. Natl. Acad. Sci. USA.* 106, 6597–6602. doi: 10.1073/pnas.0812832106.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

toxins

MDPI

Review

Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems

Nathalie Goeders, Ray Chai, Bihe Chen, Andrew Day and George P. C. Salmond *

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK; ng394@cam.ac.uk (N.G.); rc636@cam.ac.uk (R.C.); bc407@cam.ac.uk (B.C.); awd33@cam.ac.uk (A.D.)
* Correspondence: gpcs2@cam.ac.uk; Tel.: +44-1223-333-650

Academic Editor: Anton Meinhart
Received: 16 August 2016; Accepted: 19 September 2016; Published: 28 September 2016

Abstract: Toxin-antitoxin (TA) systems are small genetic modules that encode a toxin (that targets an essential cellular process) and an antitoxin that neutralises or suppresses the deleterious effect of the toxin. Based on the molecular nature of the toxin and antitoxin components, TA systems are categorised into different types. Type III TA systems, the focus of this review, are composed of a toxic endoribonuclease neutralised by a non-coding RNA antitoxin in a pseudoknotted configuration. Bioinformatic analysis shows that the Type III systems can be classified into subtypes. These TA systems were originally discovered through a phage resistance phenotype arising due to a process akin to an altruistic suicide; the phenomenon of abortive infection. Some Type III TA systems are bifunctional and can stabilise plasmids during vegetative growth and sporulation. Features particular to Type III systems are explored here, emphasising some of the characteristics of the RNA antitoxin and how these may affect the co-evolutionary relationship between toxins and cognate antitoxins in their quaternary structures. Finally, an updated analysis of the distribution and diversity of these systems are presented and discussed.

Keywords: abortive infection; altruistic suicide; type III toxin-antitoxin; bacteriophages; quaternary structures; co-evolution; pseudoknotted RNA; endoribonuclease

1. Introduction

Toxin-antitoxin (TA) systems are composed of a bacteriostatic or bactericidal toxin and a cognate antidote which is referred to as the antitoxin. In most cases, the antitoxin directly interacts with either the proteinaceous toxin or its mRNA and thus antagonises the deleterious effect of the toxin. In addition to their physical interdependence, they are linked at the genetic level and are often encoded in bicistronic operons with a promoter-proximal antitoxin gene. Save for a few Type I systems, all TA systems are encoded in operons wherein the antitoxin gene is usually found upstream of the toxin-encoding ORF. Nonetheless a few Type II systems such as HigBA and HicAB are exceptions and have a reverse genetic organisation [1]. This transcriptional organisation favours an excess of antitoxin in homeostatic conditions where the toxin is inhibited. The harmful activities of the toxins are due to their interference with essential cellular processes including DNA replication, translation, cell wall synthesis, and maintenance of membrane integrity [2–6].

The active toxin always interacts with its targets as a protein while the nature of the potent antitoxin is either RNA or protein. In addition, antitoxins neutralise their cognate toxins at several levels and act via distinct mechanisms. The nature of the antitoxin and its mode of action underpin the classification of TA systems into five Types (I–V) [7–9]. RNA antitoxins of Type I and III interact with the toxin transcripts (RNA-RNA interactions) or with the toxic protein (RNA-protein interactions) respectively. Most Type I antitoxin RNAs bind the toxin transcript in its 5'UTR region. Formation of this RNA-RNA duplex has two main effects [10]. Firstly, translation of the mRNA into the toxic protein

is hindered as the antitoxin RNA is usually complementary to the region containing the ribosome binding site (RBS) of the toxin transcript or directly competes with ribosomes. In addition to blocking translation initiation, the antitoxin/toxin RNA duplexes are the targets of cellular RNases and thus antitoxin binding ultimately leads to degradation of the toxin transcripts [6]. Aside from Type I TA systems, inhibition of toxin translation is also used by the only Type V system identified so far [7]. In this case, toxin production is turned off directly by the antitoxin which is an RNase that degrades toxin mRNAs and thus directly regulates toxin transcript levels [7]. In Type II, III, and IV systems, translation of the toxin is not directly affected. Type II systems use direct protein-protein interactions where antitoxins either mask the toxin active site or sterically hinder the toxins from reaching their target [1–3]. Type IV antitoxins also prevent toxin-target interactions but achieve this by competing with the toxins for their target, without direct contacts with their cognate toxins [8,9]. Finally, Type III TA systems fall into the category of RNA-protein interactions in which the toxin active site is occluded by the antitoxin RNA [11–14].

TA systems were originally discovered in the late 80s [15,16] and, for a long period, were defined in only two types (Type I and II) which have been extensively studied. In contrast, the archetypal Type III TA system and, subsequently, the Type IV and V systems, were discovered only recently. While data are still comparatively limited for the newer Type IV and V systems, a more holistic image is beginning to emerge regarding Type III systems, with accumulating biochemical, structural, and functional data. This review covers these systems, describing their diversity and toxin-antitoxin/abortive infection bifunctionality and discusses their impact on bacteria-phage co-evolution, given their anti-phage activity.

2. Type III TA Systems Are Split into Three Families Which Share the Same Genetic Organisation

The novelty of Type III TA systems involves the nature of RNA-protein interactions between their components [11,12]. These interactions are unusual as the toxin is involved in processing the antitoxin into its active form. More precisely, the antitoxin, a small non-coding RNA (sRNA) composed of several repeats of short nucleotide sequences, is processed into monomers of these repeats by the toxin. The heteromeric complexes adopted by Type III systems during homeostatic conditions are composed of alternating interactions between antitoxin and toxin monomers [12–14]. As for other Types of TA systems, the antitoxin has a shorter half-life than the toxin [11] but the specific details of antitoxin degradation are not yet completely clear.

At the genetic level, Type III TA systems are organised in characteristic bicistronic operons transcribed from a single promoter. The downstream toxin gene is preceded by a Rho-independent terminator that separates it from the upstream short repetitive nucleotide sequences that encode the antitoxin sRNA (Figure 1) [11,17]. Presumably, organisation in operons ensures a higher synthesis of the antitoxin compared to the toxin and thus avoids physiologically precocious, and potentially lethal, toxin activity. Unique to Type III systems toxin expression is further modulated by the presence of the inter-gene Rho-independent terminator. A final regulatory fail-safe may reside in the fact that one antitoxin sRNA will be processed into several monomers that could neutralise two or three toxins, thus further ensuring an appropriately regulated antitoxin:toxin stoichiometry.

Although all Type III TA systems share the same genetic arrangement, they can be further differentiated into three families which are classified according to the amino acid sequence similarities that they share [18]. The subfamilies are called ToxIN, CptIN, and TenpIN where the "I" and "N" represent the antitoxin and toxin components respectively. Thus, for the ToxIN system of *Pectobacterium atrosepticum* the antitoxin is referred to as $ToxI_{Pa}$, the toxin as $ToxN_{Pa}$ and both components as $ToxIN_{Pa}$ [18]. CptIN was named after the *Coprococcus catus* GD/7 system (*Co*Type III Inhibitor/toxi*N*) and the third family, TenpIN for Type III *EN*dogenous to *Photorhabdus* Inhibitor/tox*IN* [18]. While the toxin sequence directly influences the subgroup to which a particular system belongs, it is also interesting to note how their cognate antitoxins differ between and within the subgroups.

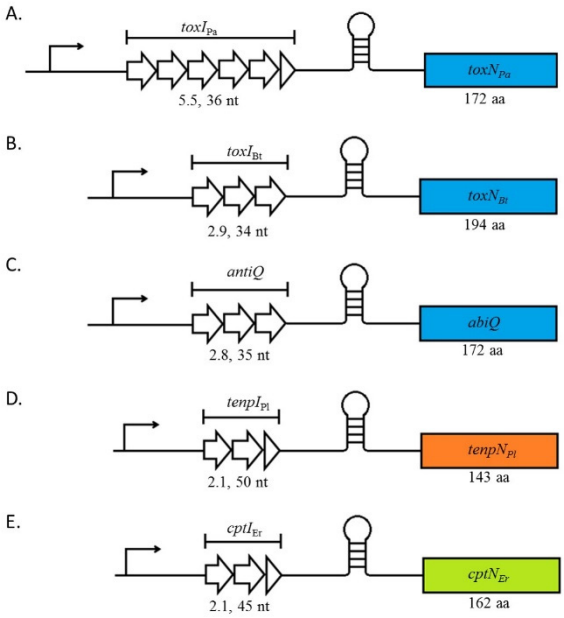


Figure 1. Genetic organisation of Type III systems. Type III systems are arranged with the antitoxin gene preceding that of the toxin—separated by a Rho-independent terminator. Five Type III TA loci are shown: (**A**) *toxIN*$_{Pa}$ located on pECA1039 from *Pectobacterium atrosepticum*; (**B**) *toxIN*$_{Bt}$ located on pAW63 from *Bacillus thuringiensis*; (**C**) AbiQ located on pSRQ900 from *Lactococcus lactis*; (**D**) *tenpIN*$_{Pl}$ from the chromosome of *Photorhabdus luminescens*; and finally (**E**) *cptIN*$_{Er}$ from the chromosome of *Eubacterium rectale*.

3. Antitoxin Length Is Important for Type III System Functions

Antitoxin repeats are a key feature of Type III systems. The number of repeats varies between systems and they have been shown to be crucial for antitoxin activity. For instance, the antitoxins of the ToxIN$_{Pa}$, ToxIN$_{Bt}$, and AbiQ systems that belong to the ToxIN family, all diverge at the primary sequence level and number of repeats, while the length of the monomers is quite conserved. ToxI$_{Pa}$, ToxI$_{Bt}$, and *antiQ* sRNAs are composed of, respectively, 36 nucleotides repeated 5.5 times, 34 nucleotides repeated 2.9 times, and 35 nucleotides repeated 2.8 times (Figure 1). In vitro, the antitoxin activity can be retained despite increasing or decreasing repeat numbers. However, the range of repeats in which each antitoxin remains functional varies. For instance, 2.5 repeats from 5.5 were necessary and sufficient for ToxI$_{Pa}$ antitoxin to inhibit its toxin [19] while at least 1.8 repeats from 2.8 were essential for the antitoxin activity of *antiQ* [17]. *antiQ* mutants containing 1.8 and 3.8 repeats were readily obtained while clones with only 0.8 of a basic repeat were inviable, suggesting that an incomplete repeat sequence is insufficient to avoid toxicity of AbiQ [17]. In addition to its TA function, the AbiQ system also acts as an abortive infection system against some phages (See below, Section 6.1). This activity is also affected by the number of *antiQ* repeats however the anti-phage activity of the system is altered independently of its toxin neutralising activity. For instance, deletion or addition of one repeat to *antiQ* decreased the phage resistance provided by the AbiQ system, indicating that the length of the wild-type *antiQ* is critical for optimal anti-phage activity. Similarly, mutations in key residues for antitoxin processing led to significant loss of anti-phage activity while a point mutation that affects pseudoknot structure increased anti-phage activity, but did not affect bacterial fitness [17].

4. Assembly of the Toxin-Antitoxin Complexes

When the paradigmatic ToxIN$_{Pa}$ system was first discovered, the activity of the toxin component was unknown and mining structural databases with the predicted structure of ToxN$_{Pa}$ gave no meaningful results [11]. Insight into its activity was gained later with the resolution of its crystal

structure and the discovery of the triangular architecture adopted by the three toxin-antitoxin monomers [12]. Resolution of the quaternary structures of further Type III systems showed that this interesting feature of Type III TAs exhibits some variations on a theme where toxin and antitoxin monomers alternate (in hexameric or tetrameric complexes) in which only RNA-protein interactions occur. A hallmark shared by all the structures is that it is the antitoxin processing that leads to the inactive, stable TA complex [12–14]. So far, the core architecture of Type III systems seems to be subfamily specific and likely depends on the length and fold of the antitoxin monomers.

4.1. The ToxIN Systems Form Triangular Heterohexamers

Most of the structural data currently available concerns the ToxIN subfamily. The quaternary structure of the ToxIN$_{Pa}$ and ToxIN$_{Bt}$ systems has been resolved (Figure 2A,B) and bioinformatic analyses predict that the AbiQ system shares the same quaternary architecture [12,13,20]. These crystal structures provided important insights into the mechanism of RNA anti-toxicity.

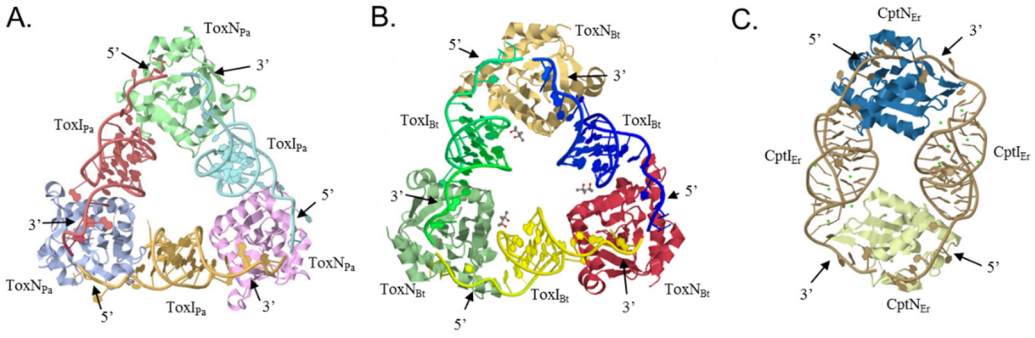


Figure 2. Crystal structures of Type III TA systems. (**A**) ToxIN$_{Pa}$ (PDB ID: 2XDD) and (**B**) ToxIN$_{Bt}$ (PDB ID: 4ATO) form heterohexameric complexes [12,13]; (**C**) CptIN$_{Er}$ (PDB ID: 4RMO) assembles into a heterotetrameric complex [14].

Both the ToxIN$_{Pa}$ and ToxIN$_{Bt}$ systems assemble into heterohexameric complexes (Figure 2A,B) that adopt a triangular architecture where toxins occupy the apices and are held together by processed antitoxin monomers [12,13]. In the complexes, each pseudoknotted ToxI RNA is bound head to tail to two separate ToxN monomers and occludes their active sites. Since ToxN$_{Pa}$ and ToxN$_{Bt}$ are both endoribonucleases, cleavage of cognate tandem RNAs into single repeats and assembly into the ToxIN complexes is likely associated with the inhibition of toxicity. Experimental data show that, while ToxN$_{Pa}$ is neutralized by both the processed and precursor forms of ToxI$_{Pa}$, the precursor ToxI$_{Pa}$ is the preferred substrate, indicating the involvement of the antitoxin processing in toxin inhibition [13]. Additionally, the pseudocontinuous arrangement of ToxI units in these structures and the 2′-3′ cyclic phosphates at the 3′ end of the processed antitoxins support this mechanism [12,13]. Finally, ToxIN$_{Pa}$ complexes can self-assemble in vitro from ToxN$_{Pa}$ combined either with processed or precursor ToxI$_{Pa}$ RNAs indicating that the self-assembly of these structures is largely mediated by the antitoxin RNAs and does not require any cellular factors or exogenous energy [13].

Both ToxI$_{Pa}$ and ToxI$_{Bt}$ antitoxin monomers fold into a classic H-type pseudoknot structure flanked by two single-stranded tails. These ends and the adjacent areas of the pseudoknot interact with their respective toxins to stabilise the trimeric structure (Figure 2A,B) [12,13]. Analysis of the crystal structure of ToxIN$_{Pa}$ showed that each ToxI$_{Pa}$ tail interacts with a different ToxN$_{Pa}$ monomer via electropositive grooves where hydrogen bonds occur between the protein and the RNA bases. The antitoxin 3′ end tail containing the 2′-3′ cyclic phosphate is held in place in this groove by five side chains, Tyr29, Lys33, Thr52, Ser53, and Lys55, that form the ToxN$_{Pa}$ active site [12]. Similarly, key areas of ToxI$_{Bt}$ that interact with ToxN$_{Bt}$ are the single stranded tails, where C19 and G20 interact with Lys148 of the toxin, and the 5′ tail of ToxI$_{Bt}$ and U10 interact with a hydrophobic pocket of the ToxN$_{Bt}$ interacting with the 3′ end [13].

4.2. CptIN$_{Er}$ Assembles into Heterotetramers

In contrast to the ToxIN systems, CptIN$_{Er}$—currently the only other Type III TA system with a solved crystal structure—assembles into tetramers (Figure 2C). Its quaternary structure is composed of two toxin monomers joined by two antitoxin RNAs [14]. The CptI monomers are longer than the corresponding examples in the ToxI repeats and the fold they adopt probably accounts for the difference in quaternary structures—with the nature of the pseudoknotted RNA perhaps driving the evolution of the toxin-antitoxin complexes.

4.3. Type III Antitoxins Form Pseudoknots

All the antitoxins crystallised so far adopt a pseudoknotted fold in their TA complexes. Pseudoknots are a recurrent RNA structural motif in which a loop forms interactions with distal bases outside the loop to form triple stranded structures. Usually the third nucleotide is an adenine enabling A-minor interactions [21]. In the quaternary structures of the TA systems, the antitoxin pseudoknots exhibit three distinct regions that interact with each other through duplex and triplex hydrogen bonds (Figure 3) [12,13]. An important feature is the core of the pseudoknot which contains three internal base triplexes. One of these internal base triplexes (triplex 3, GUU) separates the two based-paired stems of the pseudoknot with interdigitation of a guanine (Figure 3A,B). While key components are conserved in both ToxI pseudoknots, this curious structural aspect was not predicted due to the low sequence homology between the antitoxins. Despite overall similar structures, the RNA-protein interfaces show substantial differences as highlighted by the selective inhibition displayed by both antitoxins and the absence of functional cross-talk between antitoxins and toxins of the *Pectobacterium* and *Bacillus* systems [13].

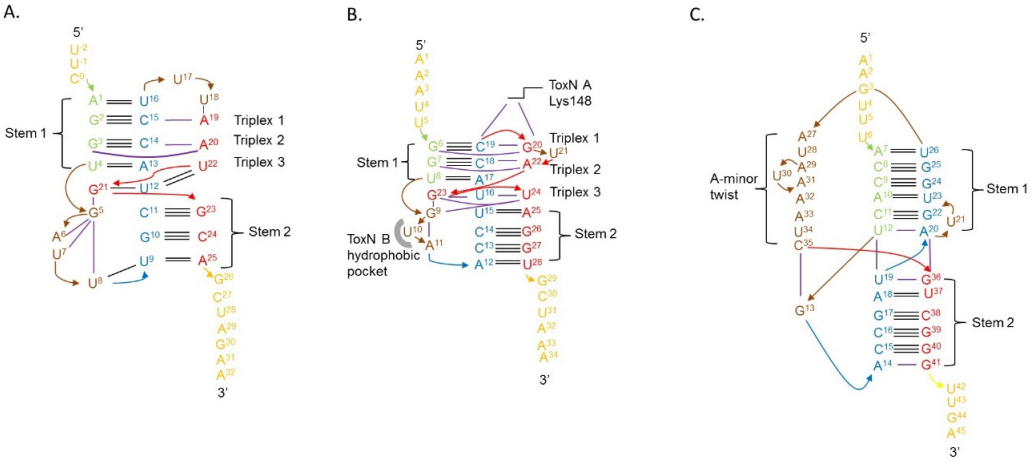


Figure 3. Pseudoknot arrangements of Type III antitoxins. (**A**) ToxI$_{Pa}$; (**B**) ToxI$_{Bt}$; and (**C**) CptI$_{Er}$. Base-base hydrogen bonds are shown by black lines. Nucleotides involved in loops are indicated in brown. Corresponding areas for each antitoxin are highlighted by similar colors.

In contrast, CptI$_{Er}$ monomers fold into an H-type pseudoknot distinctly different from the fold of the antitoxins of the ToxIN family. Firstly, CptI$_{Er}$ has two coaxial stems assembled entirely from duplex base pairing (Figure 3C) as opposed to the triplex base pairing seen in the other two antitoxins [14]. Even among RNA pseudoknots, the loops of the CptI$_{Er}$ pseudoknots are unusual. Indeed, loop 1 (L1) interacts with stem 2 (S2) while loop 2 (L2) interacts with stem 1 (S1), a configuration that leads often to triplex base pairing [21]. These canonical pseudoknot interactions are found in the ToxI RNAs where two triplex base pairs occur from the interaction of L2 with S1, although there is no triplex pairing with L1 and S2 (Figure 3A,B) [12,13]. The CptI$_{Er}$ pseudoknot is more atypical as L1 is extremely short and only consists of a single base that does not interact with S2 and is instead held in place by interactions with the 3′ end of L2 (Figure 3C) [14]. L2 of CptI$_{Er}$ is adenine rich and longer than in the

other antitoxins and forms a novel counter-clockwise A-minor twist motif [14]. This conformation appears to be necessary for its antitoxin activity, as suggested by experiments on substitution and deletion mutations that disrupted key features of this twist [14]. As with the ToxI-type antitoxins, functional cross inhibition experiments have confirmed that CptI antitoxins also show high toxin specificity [14].

5. Type III Toxins Share a Common Fold and Activity

To date, the structures of four Type III toxins have been solved; three of which belong to members of the ToxIN family (ToxIN$_{Pa}$, ToxIN$_{Bt}$ and AbiQ) and, more recently, a member of the CptIN family. All toxins share a globular β core surrounded by α-helices and loops. While the core structure is conserved, most variations are found on the surface. These variations are thought to account for cleavage and antitoxin specificity of the toxins. All Type III toxins tested so far for their mechanistic action have been shown to be endoRNases that cleave mRNAs in adenine-rich regions but with slightly different substrate specificities: ToxN$_{Pa}$ AA/AU, ToxN$_{Bt}$ A/AAAA, and AbiQ A/AAA [13,20]. As for other toxins that inhibit translation, Type III toxins initially have a bacteriostatic effect on growth [3,11,20] that ultimately leads to lethality.

5.1. The ToxN Family

ToxN$_{Pa}$ and ToxN$_{Bt}$ have the same highly twisted core of six anti-parallel β-sheets and five main variable loop regions (Figure 4A,B) [12,13]. Four of these loops, especially the long kinked helix H3, act as the major sites of interaction between the toxins and their respective antitoxins while the fifth encompasses the active site. In line with the structural data, bioinformatics analysis of the sequences of the other members of the ToxN family showed that the core fold is conserved while most sequence variability clusters in regions corresponding to these five loops [12]. The structure of AbiQ also encompasses a core of six-stranded anti-parallel β-sheets surrounded by 6 α-helices (Figure 4C) [20]. Surprisingly, the RNase activity of AbiQ was only eliminated by one site-directed mutation in Ser51Leu which is thought to make the nucleophilic attack during the cleavage [20]. The RNase activity of AbiQ was retained when replacing the Ser51 by a threonine, the equivalent residue found in the active site of ToxN [12,13,20].

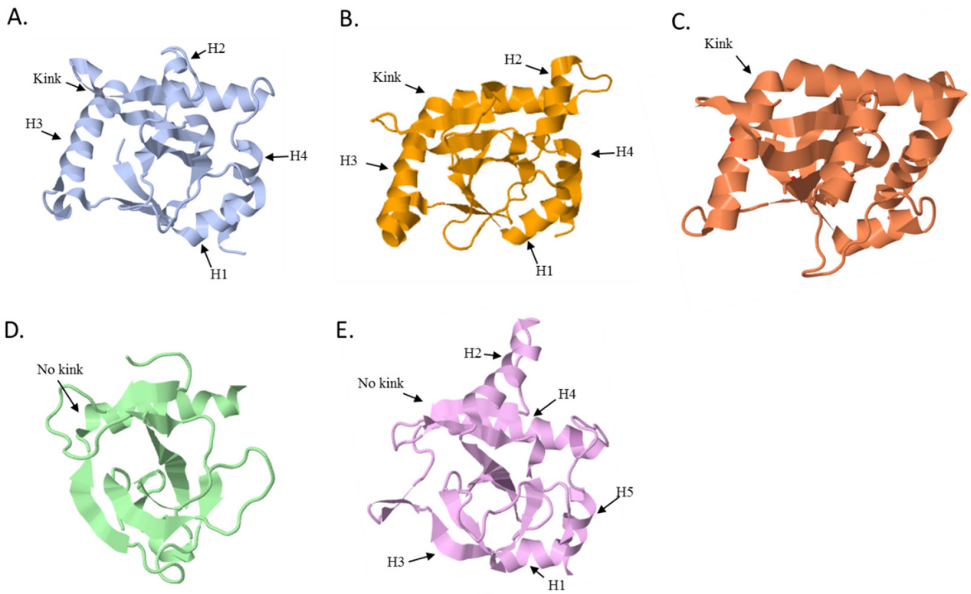


Figure 4. Structures of the toxins. (**A**) Structure of ToxN$_{Pa}$ (PDB ID: 2XDD, [12]); (**B**) ToxN$_{Bt}$ (PDB ID: 4ATO, [13]); (**C**) AbiQ (PDB ID: 4GLK, [20]); (**D**) Kid (PDB ID: 1M1F, [22]); and (**E**) CptN$_{Er}$ (PDB ID: 4RMO, [14]). The presence or absence of the kink in helix 3 (H3) is indicated for each toxin.

5.2. The CptN Family

$CptN_{Er}$ shares the same core fold found in the other toxins despite its lower primary sequence identity with members of the ToxN family (Figure 4E). In $CptN_{Er}$, the highly twisted anti-parallel β-sheets that form the core are surrounded by four α-helices that make extensive interactions with the cognate antitoxin sRNA [14]. Compared to the ToxN family, in which $ToxN_{Pa}$, $ToxN_{Bt}$, and AbiQ have a kink in helix three [12,13,20] the equivalent helix in $CptN_{Er}$, helix H4, is much shorter [14]. However many of the hydrogen bonding networks in the RNase active site are also conserved suggesting that this toxin is also an endoRNase.

5.3. Type III Toxins Share Homology with Type II Toxins

Unexpectedly, the structure of Type III toxins shows significant homology with toxins from the Type II MazF/Kid/CcdB family (Figure 4). Members of this Type II family share a similar fold despite low sequence similarity and different mechanisms of action. For instance, Kid and MazF toxins inhibit translation by acting as endoRNases while CcdB toxins interact with DNA gyrase and affect DNA replication [23–25]. Type II and Type III toxins share very low sequence similarity, e.g., only 11% identity between $ToxN_{Pa}$ and Kid, but their respective structures encompass a similar β core fold region surrounded by helices and loops. Interestingly, despite sharing the same overall structure and the same molecular activity i.e., endoRNases, the active sites of ToxN and Kid do not match well [11,25]. Less surprisingly, differences are also found in the regions that interact with their respective antitoxins. For instance, the helix 3 region of $ToxN_{Pa}$ overlays with the Kid structure but is greatly extended in the N-terminal section of the $ToxN_{Pa}$, the main site of ToxI recognition. The equivalent helix in $CptN_{Er}$, helix H4, is much shorter and, as such, resembles more closely to the Kid toxin (Figure 4A,D,E).

6. Functions of Type III TA System

Two functions have been ascribed to Type III TA systems so far. Historically, the first function was their abortive infection (Abi) activity that protects bacterial populations from invading bacteriophages. Their second known function involves their role in stabilisation of plasmids.

6.1. Abortive Infection

Some Type III systems are active protagonists in the "molecular arms race" that occurs between bacteria and phages. As these outnumber bacteria 10-fold in some environments, the prokaryotic hosts are potentially subject to constant viral predation [26]. In response to this acute selective pressure on microbial communities, bacteria have evolved a large array of defense mechanisms to circumvent the lethal impact of viral infections [27,28].

At least some of these Type III systems are bifunctional and can act as abortive infection (Abi) systems in addition to their TA functionality. This bifunctionality is possible as both activities require the involvement of a self-poisoning protein. Abi mechanisms rely on a toxic protein that can be viewed, from an evolutionary perspective, as inducing the altruistic suicide of infected cells to prevent phage propagation in bacterial populations. Even though 23 different Abi systems have been reported, mostly carried on lactococcal plasmids, there is still a paucity of information on the molecular mechanisms involved in their function, with the exception of TA/Abi systems for which some insight has been acquired recently [29]. Abi systems have been shown to act at different steps of the phage replication cycle i.e., from DNA replication to the lysis step—but ultimately they all lead to the death of the infected bacteria [29]. TA/Abi bifunctionality is not restricted to Type III systems as TA systems from Type I to IV have been shown to protect bacteria from bacteriophages [8,11,30–32]. While molecular details are still elusive in most cases, a working model is that—following toxin activation by the infecting phage—dissemination of phage progeny in the bacterial population is restricted. Toxin activation could rely on the differential stability of the antitoxin and toxin products.

Post-infection, this characteristic of TA systems could favour a state of free toxins upon the depletion of the shorter lived antitoxins.

The ToxIN$_{Pa}$ system of *P. atrosepticum* was identified originally through its phage resistance capacity encoded by a cryptic plasmid. Subcloning of different plasmid fragments eventually narrowed the anti-phage activity down to the ToxIN system genetic module [11]. ToxIN$_{Pa}$ can inhibit a spectrum of phages infecting multiple Gram-negative enterobacteria such as *P. atrosepticum*, *Escherichia coli*, and *Serratia* spp. ([10] and unpublished data). This dual Abi and TA functionality was shown later to be shared by some other Type III TA systems. The chromosomal *tenpIN* locus from *Photorhabdus luminescens*, effectively aborts environmental coliphages when expressed from a plasmid in *E. coli* [11,18,33] and, likewise, AbiQ is effective, in Gram-positive *Lactococcus lactis*, against members of the common 936 and c2 phage groups as well as rarer lactococcal phages. When expressed in *E. coli*, AbiQ also protects this Gram-negative host from some coliphages [33]. Therefore, the protective effect against phages appears to be independent of the organism in which the systems are present [11,18,33]. Phage resistance may be also independent of the original genetic location as both plasmid and chromosomal systems can effectively inhibit phage replication [11,18]. Similarly, no correlation between the Type III TA systems families and Abi activity can be drawn currently. For instance, despite proficient anti-phage activity of the ToxIN$_{Pa}$ and AbiQ systems, no anti-phage activity could be shown for the closely related ToxIN$_{Bt}$ system in its native host [13] and while both phenotypes have been observed for members of the ToxIN family, no anti-phage activity has yet been observed for the two tested CptIN systems from *Eubacterium rectale* and *Ruminococcus torque* [14,18].

A further aspect of anti-phage activity of Type III systems is the specificity these systems show against subgroups of sensitive phages [11,18]. No correlation has been found between phage families (*Myoviridae*, *Siphoviridae*, and *Podoviridae*) and abortive infection [33,34]. It is not known whether phages that are not aborted by the TA/Abi systems have naturally and actively evolved resistance mechanisms against these systems in the perpetual arms race that occurs between them and their bacterial hosts, or whether they "simply" do not activate the toxin of these systems. Some examples of mechanisms selected by phages to avoid Abi systems in general (as well as TA/Abi systems) are known in the literature [34]. For instance, point mutations in key phage products have been identified as the basis of escape mechanisms. More recently it was shown that phage T4 has an ADP ribosyltransferase that chemically modifies the Type II MazF toxin to downregulate its activity and thus avoid this TA/Abi system [32]. Another phage escape mechanism includes two examples where the phages evolved mimics of the toxin's substrates. In the case of the Type II RnlAB/LsoAB systems of *E. coli* and phage T4, the phage uses the Dmd protein which acts like a Type II antitoxin and directly interacts with the toxins. In contrast to the canonical antitoxins of RnlA and LsoA, Dmd is able to cross-neutralise several toxins. Recent crystallographic studies showed that Dmd is thought to have a different inhibition mechanism and directly interacts with the toxin active site by mimicking the toxin substrates [35].

In the Type III systems, the ToxIN-sensitive phage TE was shown to produce low frequency spontaneous mutants that escaped the ToxIN$_{Pa}$ system [36]. Analysis of the escape mutants revealed that the majority of the mutants had extended a short viral sequence similar to the repeats of the ToxI sRNA into a 'pseudo-ToxI' which functionally suppressed the toxin. In one case, recombination had allowed the phage escape mutant to obtain natural ToxI repeats from the original plasmid antitoxin sequences. Both scenarios allowed phage replication unaffected by the Abi/TA system [36].

The precise molecular mechanisms of these bifunctional TA/Abi systems are still under investigation. In agreement with the current model, the toxin endoRNase activity and the Abi phenotype have been shown to be linked [19]. However, this may not be a universal situation because mutagenesis of key toxin residues can lead to the loss of the Abi phenotype despite retained RNase activity [20] indicating that the details of the toxin activation are more subtle. This leads to the question as to how Type III TA/Abi systems are actually activated and the molecular basis of the differential sensitivity of different phages. Phage products that are involved in the activation of Type III

TA/Abi systems might directly interact with the toxin, the antitoxin, or both, to prevent assembly of the TA heterocomplexes, or to disrupt them. Alternatively, a phage product might interrupt transcription of the TA operon or affect antitoxin RNA stability, thus causing imbalance in the TA components. Both scenarios would lead to an excess of free toxin, which could inhibit mRNA translation, leading to cell growth arrest and thereby ultimately prevent the release of phage progeny. The nature of the phage products involved and the way they interact (directly or indirectly) with the TA system is unknown. Experimental data for the ToxIN$_{Pa}$ and the AbiQ systems showed that toxin levels are not affected during phage infections nor is transcription of the TA systems increased [19,20]. However transcripts of the AbiQ system, expressed constitutively before infection, decrease during the infection, possibly as part of a general infection phenomenon [20]. Based on experimental data, it is thought that the TA systems are activated at late steps of phage replication, at least in the AbiQ system. It has been shown that phage DNA is replicated in infected cells in the presence of the AbiQ system, as indicated by the accumulation of concatemeric viral DNA [37].

6.2. Plasmid Inheritance through Addiction

Another function known for Type III systems is plasmid addiction. In contrast to the Abi phenomenon where TA systems protect bacteria from invading DNA, when acting as "addiction" modules they ensure the stable inheritance of plasmids in bacterial populations.

This feature of TA systems also relies on the antagonism of the self-poisoning essence of the toxin and the labile nature of the antitoxin [11]. It supposes a continuous synthesis of the antitoxin in order to keep the toxin in an inhibited state and thus renders the cell 'addicted' to the TA system. Consequently, when present on mobile genetic elements such as plasmids, these "addiction" modules promote the maintenance of the DNA molecule encoding the TA system and ensure the continued presence of plasmids in bacterial populations via killing of plasmid-free segregant cells by the toxins [15,38]. Historically, this mechanism has also been named post-segregational killing (PSK) as lethality arises when plasmids are not segregated into both daughter cells during cell division [38]. Alternatively, given the initially bacteriostatic—rather than bactericidal—nature of Type III toxins [11,20], mis-segregation could lead to a transitory growth inhibition. Bacteria that retained the plasmid would then be able to outgrow those that "lost" it, thus ensuring plasmid maintenance at the population level.

Plasmid addiction is a function found in many TA systems of Type I and II and has also been shown for some Type III TA systems. For example, ToxIN$_{Pa}$ and CptIN$_{Er}$ increased plasmid retention in *E. coli* W3110 to 100%, compared to 50% loss of the control vector [13,14]. Notably, the fact that CptIN$_{Er}$ is located on the chromosome of the original host, yet could promote plasmid maintenance when cloned, suggests that CptIN$_{Er}$ can promote its own retention and might be disseminated through horizontal gene transfer [14].

In addition to the post-segregational killing during the exponential growth phase, one Type III system has been show to ensure plasmid maintenance in stress conditions leading to sporulation of *Bacillus subtilis*. During sporulation, bacteria differentiate into spores, a metabolically dormant cell-type that can survive adverse environments and switch back to vegetative growth in favorable conditions. The frequency of plasmid loss during sporulation is higher than in vegetative growth [39] and the rate of plasmid loss can vary from 5% to 95% in *Bacillus* [40]. The *B. subtilis* ToxIN$_{Bt}$ system favours plasmid retention in this specific context by decreasing plasmid loss around 10-fold (from 58% for the control plasmid to 6% when the TA system was present) [39]. This was achieved by reducing the proportion of cells that form mature spores—probably by the action of the toxin in plasmid-free forespores. Only forespores that inherited a plasmid copy were able to mature and so this mechanism ensures plasmid maintenance throughout the environmental stress that precipitated sporulation. It is formally possible that other types of TA systems might affect plasmid retention by similar routes during sporulation [39].

7. New Bioinformatic Analysis Reveals a Significant Increase in Potential Novel Systems

Mining databases for ORFs and subsequent analysis of the proteins encoded by them can be a powerful tool to gain evolutionary insight into genes encoding TA systems [41,42]. Investigating the occurrence of Type III systems was initially complicated as bioinformatic analysis of non-coding RNAs is still problematic. Solving the structure of ToxN made it possible to identify new members of the Type III TA family by performing de novo structure-based homology searches [18]. A set of criteria was used to determine which ToxN structural homologues found by FUGUE [43], a program for sequence-structure comparison, are truly from Type III TA systems. The logic of the definition was as follows: the ToxN homologues should be preceded by a palindromic repeat acting as a Rho-independent terminator, which in turn should be preceded by sequences with a similar organization to *toxI*—composed of a tandem array of nucleotide repeats. Subsequently, samples from each family were taken and exhaustive BLASTp searches were performed, yielding a final list of 125 putative type III TA systems. The majority of hits were found on bacterial chromosomes and plasmids of Firmicutes, Fusobacteria, and Proteobacteria as well as Archaea. One *toxIN* locus was also encoded on a prophage [18]. Using these criteria, 37 putative Type III loci were identified and were divided into three independent families according to their protein sequences as described previously [18].

Given the near exponential increase of sequenced genomes available in databases during the last few years, we did a new BLASTp search. Using the amino acid sequences of the characterized ToxN$_{Pa}$, ToxN$_{Bt}$, AbiQ, TenpN$_{Pl}$, and CptN$_{Er}$ toxins as input, we did an initial BLASTp search against the Bacteria and Viruses databases of Uniprot. We identified 603 potential Type III toxins with an *E*-value <0.001. A selection of these predicted toxins were further analysed following the previously established criteria. After validation as members of putative Type III TA systems, these toxins were used for a further round of BLASTp. The new round of BLASTp searches with the selected toxin homologues gave a few new putative toxins indicating that the search is exhaustive. The different toxin groups seem to cluster to particular phylogenetic groups. For instance, the ToxN family is primarily found in Firmicutes, Fusobacteria and occasionally other species. Systems from this group also seem associated with the mobilome as they can be found on plasmid and chromosomal locations, sometimes in the close vicinity of transposase elements. Despite the increased number of predicted systems, Type III TA systems appear largely restricted to a limited group of phyla as only 8 out of the currently estimated 74–76 bacterial phyla have at least one potential system [44]. Whether Type III TA systems are indeed only confined to a restricted number of bacteria, or whether the current distribution pattern is biased by the limited phylogenetic breadth of genome sequencing, is unknown. It is intriguing that Type III TA systems seem to be confined to a few phylogenetic groups. As the toxin mode of action, i.e., cleaving RNAs, is a general one that does not require any known external factors, and given that they are found on plasmids, one might expect them to be spread through more phylogenetic groups, in a similar way to the Type II systems where the majority of toxins are also RNases [41,42,45].

Advances in technologies such as single-cell genomics and metagenomics will enable the sequencing of uncultivated organisms from diverse habitats [44,46] and might affect the picture of TA system distribution as the current "tree of life" expands and culturing-associated biases are overcome [43,47]. When placing the consensus antitoxin repeats next to a phylogenetic tree based on the protein toxin sequences, it becomes clear that the antitoxins of closely related toxins are conserved as well (Figure 5). This consistency provided evidence for the notion of co-evolution of toxin and antitoxin components. The length and the primary sequence of the consensus repeats are usually quite conserved within the different toxin groups. On the other hand, the number of repeats is prone to slight variations.

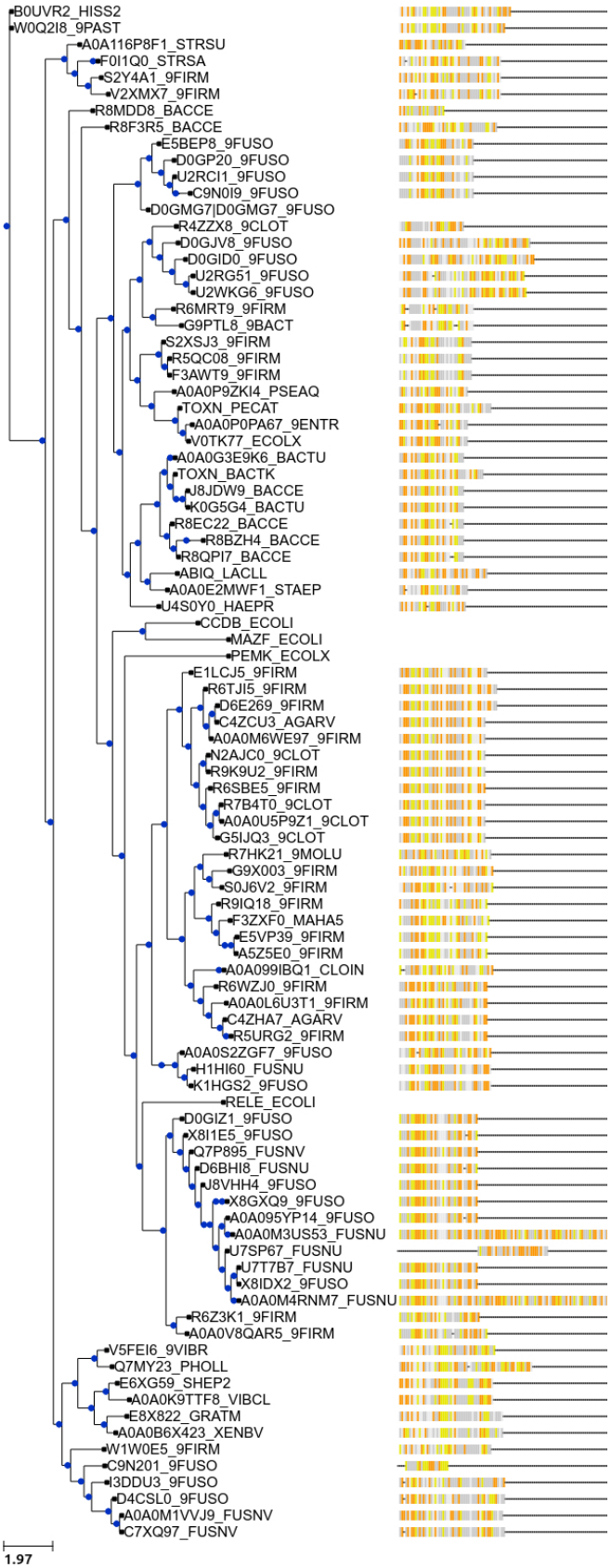


Figure 5. Phylogenetic tree based on representative toxins of putative Type III TA systems associated with the consensus repeats of the antitoxins: adenine (grey), guanine (white), cytosine (yellow), and uracil (orange). Toxin protein sequences were aligned using MUSCLE [48]. The tree has been constructed using PhyML and nearest neighbour interchange [49,50] and the figure has been made using Ete3 [51].

8. Conclusions

Work in recent years has increased our understanding of bacterial Type III TA systems. However, it is abundantly clear that our understanding of the molecular mechanisms involved in Type III systems and their activation is still rudimentary. These TA systems provide valuable reagents for fundamental biochemical studies on protein:RNA interactions and the study of quaternary nucleoprotein complex assemblies. The ecological and fitness implications of bacterial carriage of Type III TA systems have yet to be investigated. Furthermore, the role of Type III systems in the evolution, replication, and physiology of both bacteria and their viral parasites warrants deeper consideration.

Acknowledgments: Work in the Salmond lab is supported by the BBSRC, UK; N.G. was supported by the Fonds National de la Recherche Luxembourg (9118191); B.C. was supported by a Cambridge International Scholarship from the Cambridge Commonwealth, European & International Trust; and A.D. was supported by a BBSRC -DTP studentship.

Author Contributions: All authors were involved in the writing and revision of the review article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jurenaite, M.; Markuckas, A.; Sužiedeliene, E. Identification and characterization of type II toxin-antitoxin systems in the opportunistic pathogen *Acinetobacter baumannii*. *J. Bacteriol.* **2013**, *195*, 3165–3172. [CrossRef] [PubMed]
2. Yamaguchi, Y.; Inouye, M. An endogenous protein inhibitor, YjhX (TopAI), for topoisomerase I from *Escherichia coli*. *Nucleic Acids Res.* **2015**, *43*, 10387–10396. [PubMed]
3. Goeders, N.; Van Melderen, L. Toxin-antitoxin systems as multilevel interaction systems. *Toxins* **2014**, *6*, 304–324. [CrossRef] [PubMed]
4. Harms, A.; Stanger, F.V.; Scheu, P.D.; de Jong, I.G.; Goepfert, A.; Glatter, T.; Gerdes, K.; Schirmer, T.; Dehio, C. Adenylylation of gyrase and topo IV by FicT toxins disrupts bacterial DNA topology. *Cell Rep.* **2015**, *12*, 1497–1507. [CrossRef] [PubMed]
5. Sofos, N.; Xu, K.; Dedic, E.; Brodersen, D.E. Cut to the chase—Regulating translation through RNA cleavage. *Biochimie* **2015**, *114*, 10–17. [CrossRef] [PubMed]
6. Brielle, R.; Pinel-Marie, M.-L.; Felden, B. Linking bacterial type I toxins with their actions. *Curr. Opin. Microbiol.* **2016**, *30*, 114–121. [CrossRef] [PubMed]
7. Wang, X.; Lord, D.M.; Cheng, H.-Y.; Osbourne, D.O.; Hong, S.H.; Sanchez-Torres, V.; Quiroga, C.; Zheng, K.; Herrmann, T.; Peti, W.; et al. A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nat. Chem. Biol.* **2012**, *8*, 855–861. [CrossRef] [PubMed]
8. Dy, R.L.; Przybilski, R.; Semeijn, K.; Salmond, G.P.C.; Fineran, P.C. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res.* **2014**, *42*, 4590–4605. [CrossRef] [PubMed]
9. Masuda, H.; Tan, Q.; Awano, N.; Yamaguchi, Y.; Inouye, M. A novel membrane-bound toxin for cell division, CptA (YgfX), inhibits polymerization of cytoskeleton proteins, FtsZ and MreB, in *Escherichia coli*. *FEMS Microbiol. Lett.* **2012**, *328*, 174–181. [CrossRef] [PubMed]
10. Wen, J.; Fozo, E. sRNA antitoxins: More than one way to repress a toxin. *Toxins* **2014**, *6*, 2310–2335. [CrossRef] [PubMed]
11. Fineran, P.C.; Blower, T.R.; Foulds, I.J.; Humphreys, D.P.; Lilley, K.S.; Salmond, G.P. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 894–899. [CrossRef] [PubMed]
12. Blower, T.R.; Pei, X.Y.; Short, F.L.; Fineran, P.C.; Humphreys, D.P.; Luisi, B.F.; Salmond, G.P.C. A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat. Struct. Mol. Biol.* **2011**, *18*, 185–190. [CrossRef] [PubMed]
13. Short, F.L.; Pei, X.Y.; Blower, T.R.; Ong, S.-L.; Fineran, P.C.; Luisi, B.F.; Salmond, G.P.C. Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E241–E249. [CrossRef] [PubMed]

14. Rao, F.; Short, F.L.; Voss, J.E.; Blower, T.R.; Orme, A.L.; Whittaker, T.E.; Luisi, B.F.; Salmond, G.P.C. Co-evolution of quaternary organization and novel RNA tertiary interactions revealed in the crystal structure of a bacterial protein-RNA toxin-antitoxin system. *Nucleic Acids Res.* **2015**, *43*, 9529–9540. [CrossRef] [PubMed]
15. Ogura, T.; Hiraga, S. Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4784–4788. [CrossRef] [PubMed]
16. Gerdes, K.; Rasmussen, P.B.; Molin, S. Unique type of plasmid maintenance function: Postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3116–3120. [CrossRef] [PubMed]
17. Bélanger, M.; Moineau, S. Mutational analysis of the antitoxin in the lactococcal type III toxin-antitoxin system AbiQ. *Appl. Environ. Microbiol.* **2015**, *81*, 3848–3855. [CrossRef] [PubMed]
18. Blower, T.R.; Short, F.L.; Rao, F.; Mizuguchi, K.; Pei, X.Y.; Fineran, P.C.; Luisi, B.F.; Salmond, G.P.C. Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. *Nucleic Acids Res.* **2012**, *40*, 6158–6173. [CrossRef] [PubMed]
19. Blower, T.R.; Fineran, P.C.; Johnson, M.J.; Toth, I.K.; Humphreys, D.P.; Salmond, G.P.C. Mutagenesis and functional characterization of the RNA and protein components of the *toxIN* abortive infection and toxin-antitoxin locus of *Erwinia*. *J. Bacteriol.* **2009**, *191*, 6029–6039. [CrossRef] [PubMed]
20. Samson, J.E.; Spinelli, S.; Cambillau, C.; Moineau, S. Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system: AbiQ, a type III toxin-antitoxin system. *Mol. Microbiol.* **2013**, *87*, 756–768. [CrossRef] [PubMed]
21. Butcher, S.E.; Pyle, A.M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **2011**, *44*, 1302–1311. [CrossRef] [PubMed]
22. Hargreaves, D.; Santos-Sierra, S.; Giraldo, R.; Sabariegos-Jareño, R.; de la Cueva-Méndez, G.; Boelens, R.; Díaz-Orejas, R.; Rafferty, J.B. Structural and functional analysis of the Kid toxin protein from *E. coli* plasmid R1. *Structure* **2002**, *10*, 1425–1433. [CrossRef]
23. Loris, R.; Dao-Thi, M.-H.; Bahassi, E.M.; Van Melderen, L.; Poortmans, F.; Liddington, R.; Couturier, M.; Wyns, L. Crystal structure of CcdB, a topoisomerase poison from *E. coli*. *J. Mol. Biol.* **1999**, *285*, 1667–1677. [CrossRef] [PubMed]
24. Kamada, K.; Hanaoka, F.; Burley, S.K. Crystal structure of the MazE/MazF complex: Molecular bases of antidote-toxin recognition. *Mol. Cell* **2003**, *11*, 875–884. [CrossRef]
25. Diago-Navarro, E.; Hernandez-Arriaga, A.M.; López-Villarejo, J.; Muñoz-Gómez, A.J.; Kamphuis, M.B.; Boelens, R.; Lemonnier, M.; Díaz-Orejas, R. *parD* toxin-antitoxin system of plasmid R1—Basic contributions, biotechnological applications and relationships with closely-related toxin-antitoxin systems. *FEBS J.* **2010**, *15*, 3097–3117. [CrossRef] [PubMed]
26. Weinbauer, M.G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **2004**, *28*, 127–181. [CrossRef] [PubMed]
27. Labrie, S.J.; Samson, J.E.; Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **2010**, *8*, 317–327. [CrossRef] [PubMed]
28. Koskella, B.; Brockhurst, M.A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **2014**, *38*, 916–931. [CrossRef] [PubMed]
29. Chopin, M.-C.; Chopin, A.; Bidnenko, E. Phage abortive infection in lactococci: Variations on a theme. *Curr. Opin. Microbiol.* **2005**, *8*, 473–479. [CrossRef] [PubMed]
30. Pecota, D.C.; Wood, T.K. Exclusion of T4 phage by the *hok/sok* killer locus from plasmid R1. *J. Bacteriol.* **1996**, *178*, 2044–2050. [PubMed]
31. Hazan, R.; Engelberg-Kulka, H. *Escherichia coli mazEF*-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genom.* **2004**, *272*, 227–234. [CrossRef] [PubMed]
32. Alawneh, A.M.; Qi, D.; Yonesaki, T.; Otsuka, Y. An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the *Escherichia coli* MazF toxin of a toxin-antitoxin module: ADP-ribosylation of *E. coli* MazF by T4 Alt. *Mol. Microbiol.* **2016**, *99*, 188–198. [CrossRef] [PubMed]
33. Samson, J.E.; Belanger, M.; Moineau, S. Effect of the abortive infection mechanism and type III toxin/antitoxin system AbiQ on the lytic cycle of *Lactococcus lactis* Phages. *J. Bacteriol.* **2013**, *195*, 3947–3956. [CrossRef] [PubMed]

34. Blower, T.R.; Evans, T.J.; Przybilski, R.; Fineran, P.C.; Salmond, G.P.C. Viral evasion of a bacterial suicide system by RNA–based molecular mimicry enables infectious altruism. *PLoS Genet.* **2012**, *8*, e1003023. [CrossRef] [PubMed]
35. Samson, J.E.; Magadan, A.H.; Sabrie, M.; Moineau, S. Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **2013**, *11*, 675–687. [CrossRef] [PubMed]
36. Wei, Y.; Gao, Z.; Zhang, H.; Dong, Y. Structural characterizations of phage antitoxin Dmd and its interactions with bacterial toxin RnlA. *Biochem. Biophys. Res. Commun.* **2016**, *472*, 592–597. [CrossRef] [PubMed]
37. Emond, E.; Dion, E.; Walker, S.A.; Vedamuthu, E.R.; Kondo, J.K.; Moineau, S. AbiQ, an abortive infection mechanism from *Lactococcus lactis*. *Appl. Environ. Microbiol.* **1998**, *64*, 4748–4756. [PubMed]
38. Sengupta, M.; Austin, S. Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. *Infect. Immun.* **2011**, *79*, 2502–2509. [CrossRef] [PubMed]
39. Short, F.L.; Monson, R.E.; Salmond, G.P.C. A Type III protein-RNA toxin-antitoxin system from *Bacillus thuringiensis* promotes plasmid retention during spore development. *RNA Biol.* **2015**, *12*, 933–937. [CrossRef] [PubMed]
40. Turgeon, N.; Laflamme, C.; Ho, J.; Duchaine, C. Evaluation of the plasmid copy number in *B. cereus* spores, during germination, bacterial growth and sporulation using real-time PCR. *Plasmid* **2008**, *60*, 118–124. [CrossRef] [PubMed]
41. Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **2009**, *4*, 19. [CrossRef] [PubMed]
42. Leplae, R.; Geeraerts, D.; Hallez, R.; Guglielmini, J.; Dreze, P.; Van Melderen, L. Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **2011**, *39*, 5513–5525. [CrossRef] [PubMed]
43. Shi, J.; Blundell, T.L.; Mizuguchi, K. FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **2001**, *310*, 243–257. [CrossRef] [PubMed]
44. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K.; et al. A new view of the tree of life. *Nat. Microbiol.* **2016**, *1*, 16048. [CrossRef] [PubMed]
45. Guglielmini, J.; Van Melderen, L. Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. *Mob. Genet. Elem.* **2011**, *1*, 283–306. [CrossRef] [PubMed]
46. Rinke, C.; Schwientek, P.; Sczyrba, A.; Ivanova, N.N.; Anderson, I.J.; Cheng, J.-F.; Darling, A.; Malfatti, S.; Swan, B.K.; Gies, E.A.; et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **2013**, *499*, 431–437. [CrossRef] [PubMed]
47. Spang, A.; Ettema, T.J.G. Microbial diversity: The tree of life comes of age. *Nat. Microbiol.* **2016**, *1*, 16056. [CrossRef] [PubMed]
48. Eddy, S.R. What is a hidden Markov model? *Nat. Biotechnol.* **2004**, *22*, 1315–1316. [CrossRef] [PubMed]
49. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef] [PubMed]
50. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [CrossRef] [PubMed]
51. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [CrossRef] [PubMed]